

This is a repository copy of *Processing Relative Clauses Across Comprehension and Production: Similarities and Differences*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/134195/>

Version: Accepted Version

Article:

Santi, Andrea, Grillo, Nino orcid.org/0000-0002-8224-365X, Molimpakis, Emilia et al. (1 more author) (2019) *Processing Relative Clauses Across Comprehension and Production: Similarities and Differences*. *Language Cognition and Neuroscience*. pp. 170-189. ISSN: 2327-3801

<https://doi.org/10.1080/23273798.2018.1513539>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Processing Relative Clauses Across Comprehension and Production: Similarities and Differences

Andrea Santi^{1*}, Nino Grillo², Emilia Molimpakis¹, Michael Wagner³

¹*Department of Linguistics, University College London, London, UK*

²*Department of Language and Linguistic Science, University of York, York, UK*

³*Department of Linguistics, McGill University, Montreal, Canada*

*Corresponding Author:

Andrea Santi
a.santi@ucl.ac.uk
2 Wakefield Street
London WC1N 1Pf

Processing Relative Clauses Across Comprehension and Production: Similarities and Differences

We compare the processing of relative clauses in comprehension (self-paced reading) and production (planned production). We manipulated the locality of two syntactic dependencies: filler-gap (subject vs object gap) and subject-verb (center-embedded vs right-branched). The non-local filler-gap dependency resulted in a longer embedded predicate duration, across domains, consistent with memory-based accounts. For the non-local subject-verb dependency, we observe longer reading times at the main verb, but in production a greater likelihood and duration of a pause *preceding* the main verb. We argue that this result stems from the cost of computing the restriction, which manifests as a prosodic break. In the context of the subject-verb dependency manipulation, we also revisit the source of interpretation break-down in multiple center-embedding. Generally, our findings imply that memory-based accounts are adequate for filler-gap, but not subject-verb, dependencies and production studies can aid in understanding complexity effects.

Keywords: relative clause; memory; production; comprehension; syntactic dependency

Introduction

Language processing requires that we resolve syntactic relations (eg, agreement, thematic) between words that can be adjacent to one another or separated by many other words (see the words in bold within the sentences in Table 1). Psycholinguistics has been dedicated to investigating these dependencies and the complexity effects they give rise to. These effects are often explained in terms of costs arising from memory

mechanisms (Gibson, Desmet, Grodner, Watson, & Ko, 2005; Gordon, Hendrick, & Johnson, 2001; Van Dyke & McElree, 2006). Two types of dependencies have been the focus of attention: (1) filler-gap dependency, and (2) subject-verb dependency. Both of these can be manipulated for their length (ie, locality) with relative clause structures to be either long or short (see table 1). As we will see in a bit, different theories of memory postulate different “length” manipulations to increase processing costs. Previous studies investigating these dependencies for distance-based complexity effects have mostly focused on comprehension methods (Gibson et al., 2005; Gordon et al., 2001; Staub, 2010; Traxler, Morris, & Seely, 2002; Van Dyke, 2007; Van Dyke & McElree, 2006), however, a recent production study targeting sentences with a filler-gap dependency (Scontras, Badecker, Shank, Lim, & Fedorenko, 2014)¹ adds to the large body of work on the comprehension side in showing complexity effects for distance-based manipulations. The subject-verb dependency has received much less attention in comprehension studies (Grodner & Gibson, 2005; Van Dyke, 2007; Van Dyke & McElree, 2011) and no work, to our knowledge, has investigated it in production (outside of agreement attraction errors).

Here we study both dependencies across both comprehension (self-paced reading) and production (planned production) methods using identical stimuli in establishing a broader perspective of the (a)symmetry across domains. Based on current perspectives on the relationship between comprehension and production, which are backed by empirical data (Kempen, Olsthoorn, & Sprenger, 2012; Tooley & Bock, 2014), we hypothesized that memory mechanisms would be similarly recruited across the two domains. This would be observed as an *alignment* of the complexity effects

¹ See also Santi, Grillo, Grodzinsky, and Wagner (2011).

(surfacing as longer durations for longer dependency distance) at the point of dependency resolution. However, in production, the producer knows in advance what is coming, so we also expected to observe some effects *earlier* in production. For example, in the case of planning a more complex constituent we expected longer production durations. We replicate and extend earlier studies' findings of longer reading durations for longer filler-gap dependencies in comprehension and extend it to production durations. Moreover, this result converges on the same location within the sentence (ie, point of dependency resolution) in support of memory mechanisms being similarly recruited across the two domains. Interestingly, in the case of the subject-verb dependency, we see the complexity effect arise immediately prior to the point of dependency resolution in production (ie, before the verb, *ignored*) as longer pauses, but at the point of resolution (ie, at the verb, *ignored*) in comprehension. Given the complementarity of these methods, we provide a novel interpretation of the *complexity effect* observed with a lengthened subject-verb dependency (via relative clause modification) that is independent of memory-demands: closing off the subject and computing the restriction on the subject that emerges at the prosodic level as a break.

Insert Table 1 here

Relationship Between Comprehension and Production

Recent models of production and comprehension (Kempen et al., 2012; Pickering & Garrod, 2007; Pickering, McLean, & Branigan, 2013) propose that both operate on linguistic information (eg, syntax, lexicon) within a common workspace, using similar processes despite their obvious differences in inputs and outputs. Moreover, they portray production and comprehension as dependent, interacting systems. From the

early work of Levelt (1983) it has been argued that during production, the producer engages in self-monitoring through the comprehension system. Thus, it can be argued that what may be labelled a “production task” still relies to some degree on comprehension and vice versa. For example, in one version of production priming technique (Tooley & Bock, 2014), one reads and produces a prime sentence and then after a filled delay needs to read and produce a target sentence. During production, the producer must track their output (ie, comprehend it) to ensure they are accurately producing the sentence. Likewise, in self-paced reading one is likely silently producing the sentence they are reading and thereby generating a silent prosody (Fodor, 1998, 2002). Thus, labelling tasks as “production” or “comprehension” seems to stem from the relative prominence of one of these interacting systems, or more pragmatically what is directly being measured, rather than their independent investigation. This is consistent with the view that the two modalities process information in similar ways in working to *align* the linguistic representations across speaker and hearer (Garrod & Pickering, 2004). A close tie between generation and parsing mechanisms is also advocated in MacDonald (2013) and Momma & Phillips (2018).

The above claim is compatible with there also being undeniable ways in which comprehension and production differ. Tooley & Bock (2014) provide a nice summary of the empirical support for this: (1) comprehension precedes production in language acquisition (Benedict, 1979; Gertner, Fisher, & Eisengart, 2006; Tomasello, 2000), (2) we can understand sentences that we either have not produced or may not even be able to produce (Clark & Malt, 1984), (3) in terms of performance, comprehension seems an easier process (Recarte & Nunes, 2003) and (4) language pathology shows that the loss of production and comprehension do not perfectly align (see discussion within Hickok, 2010). However, it is likely these differences arise from tertiary factors (eg, differences

in the developmental timecourse of the required systems) than the actual use of linguistic information in each of these domains (Tooley & Bock, 2014).

Comprehension and Production: Syntactic Representations

The specific case of abstract syntactic representations across comprehension and production domains has been extensively investigated in the priming literature (Bock, 1986; Bock, Dell, Garnsey, Kramer, & Kubose, 2007; Pickering et al., 2013; Tooley & Bock, 2014). Early work suggested that priming of syntactic structures in the comprehension domain required that the verb be identical across prime and target (Arai, van Gompel, & Scheepers, 2007). In production, this was found not to be necessary; priming was present whether or not the verb was identical or varied across the prime and target. This would suggest a difference between domains. However, recent work has shown that when the same syntactic structures are used along with comparable priming procedures across both production and comprehension, abstract representations prime in both production and comprehension without lexical repetition of the verb. These results indicate that syntactic processes/representations are indeed comparable across these modalities (Tooley & Bock, 2014).

Relative Clause Processing and Memory Mechanisms

Given the results of Tooley and Bock (2014) one may also postulate then that the memory mechanisms used to resolve (or track) syntactic dependencies be similarly recruited across production and comprehension.

The memory mechanisms that are engaged during the resolution of syntactic dependencies have been modelled via the contribution of active maintenance and/or retrieval processes. In many cases, the models assume only one of these two types of processes as driving resolution difficulty (Frazier, 1987; Lewis & Vasishth, 2005; Van

Dyke & McElree, 2006; Wanner & Maratsos, 1978), however, some models advocate for a combination between the two (Fiebach, Schlesewsky, Lohmann, von Cramon, & Friederici, 2005; Gibson, 1998, 2000; Wagers & Phillips, 2014). Here we consider two prominent models that come from each of these classes. Under the *Cue-Based Retrieval* account (Lewis & Vasishth, 2005; Van Dyke & McElree, 2006), retrieval of the head of the dependency is triggered by a cue at the tail of the dependency. In the Dependency Locality Theory (DLT; Gibson, 1998, 2000), a *Storage* component is used for maintaining syntactic predictions during dependency resolution, in addition to an *Integration* component (ie, a component comparable to a cue-based retrieval mechanism). While integration costs will be incurred when relating a dependent to its head whether it was predicted or not, it is only predicted dependencies that incur a storage cost.

Using self-paced reading, Gibson et al (2005) tested two syntactic dimensions of relative clauses that have widely been considered in the literature to increase memory demands: (1) Filler-gap dependency distance (object vs. subject gaps) and (2) the grammatical position of the relative clause within the main sentence (subject-modifying vs. object-modifying relative clauses). We discuss each of these manipulations in turn.

Object vs Subject Gaps

Filler-gap dependencies can span multiple words, as in the case of object gaps or none in the case of subject gaps (see table 1). Consider the examples in Table 1. Gaps in object position are thought to be more complex than those in subject position because it involves a dependency between the filler (ie, *the reporter* in Table 1.) and the gap (marked by “_” in Table 1) that spans one or more words (ie, is nonlocal), in contrast to the local dependency with subject gaps. This asymmetry is one of the best-established effects in the literature on sentence processing, and supporting evidence has been found

in reading times (Gibson et al., 2005), eye fixations (Traxler et al., 2002) and comprehension and production errors in language acquisition and impairment (Friedmann, 2008; Friedmann, Belletti, & Rizzi, 2009; Garaffa & Grillo, 2008; Grillo, 2009; Grodzinsky, 2000).

Various explanations for this locality effect have been proposed, including the number of new discourse referents separating the filler and the gap (Gibson, 1998, 2000), or the degree of similarity between the filler and other constituents separating it from the gap, either in terms of their grammatical features (Grillo, 2009; Van Dyke & McElree, 2006) or their referential properties (Gordon et al., 2001; Van Dyke & McElree, 2006). The nonlocal filler-gap dependency (according to any of these measures) requires greater memory resources than the local one. This may be due to interference created by “similar” Noun Phrases intervening the dependency during cue-based retrieval at the embedded verb (ie, *attacked*), or the cost of actively maintaining the filler (*the reporter*, or some feature(s) of it), or the cost of integrating the filler at the tail of the dependency. It is also possible that a combination of active maintenance (of syntactic information of the filler) along with similarity-based interference during cue-based retrieval (of the lexical properties) over this distance contributes to this difficulty, as evidence from Wagers & Phillips (2014) suggests¹.

In comprehension, the greater difficulty of object compared to subject relative clauses is localized to the embedded verb. Longer reading times are observed at the embedded verb for object gaps than subject gaps. In production, there are two main dependent measures associated with difficulty. The most prevalent one is the time to initiate the utterance. This measure provides a window into the difficulty associated with the initial planning of an utterance. A more syntactically complex utterance (when number of words are controlled) generates longer initiation times (Ferreira, 1991).

However, it is also acknowledged that the speaker does not plan the entire utterance before speaking, but continues to plan upcoming material while speaking (Tooley & Bock, 2014). Word duration has been taken to reflect difficulty with this online planning. Scontras et al (2014) found that utterance initiation time is longer for sentences with object-gaps than subject-gaps (both in relative clauses and wh-questions). Additionally, Scontras et al. (2014) found that the duration of the filler (ie, *the reporter* in table 1) was longer in sentences with object-gaps than subject-gaps. The timepoint of these production effects differ from those found in comprehension, as at these points the locality of the dependency is still unknown in comprehension. Scontras et al. (2014) do not report production measures from the (embedded) verb region where comprehension difficulty is observed for this contrast. Thus, it is possible that one can also find effects at this point in time, as would be expected if memory mechanisms are recruited in a similar fashion during production.

This is an important prediction to test, as MacDonald et al (2016) disagree with a memory-based explanation of Scontras et al.'s (2014) results. MacDonald et al (2016) argue that the subject vs object relative contrast used by Scontras et al. (2014) has an alternative interpretation provided by a confound in their design. While the subject relative is the preferred structure to describe a compatible scene, the object relative has a competitor: the passive voice in the relative clause. Production difficulty with the “object relative” structure could then simply represent a difficulty in producing a sentence that has an alternative competitor (that the participants were overtly instructed not to use) rather than greater syntactic complexity. Scontras et al. (2017) reasonably argue against this by noting a similar initiation difficulty was observed for wh-object compared to wh-subject questions, where a web-based written production study demonstrated that the the wh-object question was by-far the preferred structure with

passives being produced only 5% of the time. Thus, the passive does not present a similar alternative structure in the wh-question experiment as in the relative clause one. Nonetheless, it would be helpful to further support this conclusion by demonstrating that the object vs subject relative clause complexity is observed with a method that does not provide ambiguity of structure (ie, Planned Production) and at a measurement point commensurate with memory demands – the embedded verb.

Subject-Verb Dependency Distance: Center-Embedding vs Right-Branching

Similarly, when the relative clause modifies the subject (ie, *Center-Embedding*) of the main clause a nonlocal dependency arises that is not present in the case of object modification (ie, *Right-Branching*). The non-local dependency in the case of Center-Embedding is between the head Noun of the subject and the main verb (ie, subject-verb agreementⁱⁱ). Thus, a widespread prediction, since Miller and Chomsky (1963), is that parsing is more difficult when a sentence contains a center-embedded (in this case, a sentence-medial relative clause which modifies the subject) modifier compared to a right-branched one (in this case, a sentence-final relative clause which modifies the object). Under DLT, center-embedding incurs a cost from two sources: (1) actively maintaining the prediction for a main verb throughout the relative clause (ie, the ‘storage’ cost) and (2) integration at the main verb, where its subject head is separated from it by the RC. According to the cue-based retrieval model (Van Dyke & McElree, 2006), center-embedding is predicted to incur a greater cost when the relative clause contains an object gap due to similarity-based interference arising from the relative clause subject (ie, *the senator*) that has features (ie, subjecthood) in common with the dependent (ie, *the reporter*) head, which is used as a retrieval cue at the verb.

Greater difficulty comprehending a center-embedded relative clause compared to a right-branching one, in the case of a single relative clause modifier, fits with the

observation that when there are two attached relative clauses, the two modificational structures dramatically contrast in terms of processing. Multiple right-branchings as in (1) can be parsed with relative ease, while multiple center-embeddings as in (2) typically lead to an essentially unparseable structure, at least in English.ⁱⁱⁱ

- (1) The scientist collaborated with the professor [who <the professor> advised the student [who <the student> copied the article].
- (2) The student [who the professor [who the scientist collaborated with <the professor>] advised <the student>] copied the article.

The additive effect of multiple interruptions of the matrix clause is usually taken to result in excessive computational load (Chomsky & Miller, 1963; Lewis & Vasishth, 2005; Miller & Chomsky, 1963, although see MacDonald, 2013 and Fodor, 2013 for alternative perspectives).

Previous experimental studies have focused on the comprehension of center-embedding. Puzzlingly, when looking at single cases of center embedding vs. right-branching, previous studies have either found no difference between the two (Baird & Koslick, 1974; Gibson & Fedorenko, 2013; Hakes, Evans, & Brannon, 1976; Holmes, 1973), or have found evidence suggesting that the (single-embedded) right-branching structure is in fact harder to process than the center-embedding structure (Gibson et al. 2005).

Several of these older studies were limited in that they confounded the attachment site of the RC with gap type, and/or were offline studies. Gibson et al. (2005) crossed the gap type and attachment site (recall Table 1), and conducted an on-line self-paced reading study, in order to resolve the issue. In addition to the expected effect of longer filler-gap dependencies (object gaps) resulting in longer reading durations of the entire relative clause (ie, *who the reporter attacked*), they found the relative clause to be read longer in right-branching compared to center-embedding structures, suggesting it is object-modifying relative clauses that are harder to process.

This effect is unexpected under any theoretical account based in memory mechanisms during sentence processing. Attributing the results to the linear position of the relative clause within the sentence (ie, the relative clause appears later in the sentence in a right-branching configuration) was rejected by the authors through showing that the final PP of the relative clause was in fact read faster in the right-branched relative clause than the center-embedded one, even though it appears linearly later in the sentence.

Instead of a memory-based explanation, Gibson et al (2005) proposed an additional factor contributing to the complexity of sentence processing, the *Information Flow* effect. The idea here is that firstly, there is a preference for “old” information to appear at the beginning of a declarative sentence (and “new” information at the end) and secondly, that restrictive relative clauses contain old information. Thus, in the case of the object-modified relative clauses studied, the placement of the relative clause is in conflict with the information flow of the utterance, and processing is more difficult than when the RC modifies the subject, causing a slow-down.

The information-flow account provides a rationale for why object-modifying relative clauses may be more difficult than subject-modifying ones, but one is still left wondering why it is then that multiple center-embeddings have such a devastating effect on the parser. Alternatively, Gibson et al. (2005) may not have observed any processing cost of the non-local dependency in center-embedding compared to right-branching due to the choice of measurement location being the entire relative clause rather than the main verb where both similarity based interference and the DLT expect a processing cost.

Current Experiments

In order to compare production and comprehension using the same syntactic structures and comparable procedures, as Tooley & Boch (2014) did in priming, we will directly

compare the temporal pattern observed in self-paced reading with the temporal profile of words and pauses in planned production. By establishing which durational effects are aligned in both comprehension and production, and which are not, we can gain additional insights into the source of a given durational effect. We use the relative clause stimuli from the previously discussed self-paced reading study (Gibson et al., 2005).

In our experiments, in addition to testing for effects on the entire relative clause, following the lead of Gibson et al. (2005), we will also study effects at the main verb (ie, *ignored*) since this is the point where integration or similarity-based interference should be highest in resolving the subject-verb dependency. Both Gordon et al. (2001) and Gibson's (2000) measurement of locality predict a longer main verb reading time for center-embedded compared to right-branching relative clauses. The number of new discourse referents (*attacked, the senator*) and referentially similar NPs (*the senator*) within the subject verb-dependency in center-embedded structures is greater than in right-branched structures, where there are none. The cue-based retrieval perspective of Van Dyke and McElree (2006) predicts an interaction effect. Specifically, a longer reading time of the main verb should be observed with center-embedded relative clauses containing an object gap than subject gap. The reason being, in the object extracted Relative Clause there is another subject (ie, *the senator*) along the dependency path that can be retrieved during cue-based retrieval at the verb (which cues for a subject) and thereby generates greater interference.

Possible Similarities in Comprehension and Production Results

If both tasks make use of the same memory mechanisms, we would expect similar temporal effects on reading times (RT) in self-paced reading and duration in production: (1) a longer relative clause duration (and more specifically a longer

embedded verb duration) for object filler-gap dependencies than subject ones, and (2) a longer main verb duration in center-embedding than right-branching.

Possible Differences in Comprehension and Production Results

We should also expect divergences in the results. A crucial difference between self-paced reading and planned production is that the participant knows beforehand what is coming up in the latter but not the former task. As with Scontras et al. (2014), we would expect to see a longer production duration of the head of the relative clause when containing an object gap than subject one, as it requires the planning of a more complex constituent. However, we would not expect this in comprehension, where the parser has no knowledge about what is coming up next.

A second type of divergence that might arise between the two tasks is in the *location* of an effect. Since the comprehender does not know beforehand the structure of the sentence, effects due to processing complexity at one point in the structure might only be seen in the reading times of a later point. For example, when reading, we might only observe the cost incurred by processing a complex constituent, which requires a longer time to integrate, after encountering the first word of the *next* constituent, when the reader can be sure that the constituent in question has been completed. Even if the actual processing cost occurs earlier, a reader might also continue to press the key in self-paced reading while still wrapping up the processing of earlier constituents, leading to spill-over effects. This would not be expected with production where the producer knows ahead of time where the complex constituent finishes.

To summarize, our study aims to establish whether two syntactic dependencies (filler-gap and subject verb), manipulated for locality, demonstrate difficulty effects that align across comprehension and production tasks and at locations commensurate with

memory-based accounts. Any differences across tasks or dependencies would be informative to our understanding of the processing required of these dependencies.

Experiment 1

Methods

Participants:

34 native monolingual British speakers participated (29.6 years old, 22 female). Two participants were excluded for low behavioural responses on the comprehension questions (<70% accuracy). All participants gave informed consent in accordance with UCL ethics and were compensated for participation.

Materials:

The stimuli were the same as those in Gibson et al. (2005) with minor adaptations for British English (see Appendix I for full list of items) and excluded the factor of NP-embedding. There were 32 items per condition in a 2extraction (subject, object) x 2embedding (center, right) design with an additional 64 filler sentences. Participants answered comprehension questions that followed each sentence to ensure participants read the stimuli carefully. In order to test for comprehension across the entire sentence, questions referred to the content of both the main and relative clause. Half of these required a “yes” response and half a “no” response (see 3 and 4 for an example). A mix of comprehension questions were used: some focused on thematic role assignment, others on additional aspects of the sentence, as to avoid participants taking on strategic processing.

(3) Item: *The reporter who the senator attacked on Tuesday ignored the president.*

(4) Question: *Did the reporter attack the senator?* (FALSE)

Procedure:

The sentences were presented in a moving window display in a Latin square design within *Linger* (<http://tedlab.mit.edu/~dr/Linger/>). All words were displayed

simultaneously on the screen as a series of dashes, each set of dashes masking a word, with gaps denoting the spaces between them. By pressing the spacebar with the thumb of their dominant hand, participants revealed the first word. Consequent spacebar presses re-masked this word with dashes, whilst also revealing the next word and so on until the sentence end^{iv}. Upon the final key press a comprehension question appeared to which participants responded by key presses (“F” for YES, “J” for NO). Each participant was first familiarised with the experimental method via a series of practise trials, which were excluded from data analysis. All experiments were performed in soundproof booths using a Dell© desktop computer running Windows Vista© Home Edition. A Razer© (Black Widow© 2014) gaming keyboard (1000Hz polling rate) was used for accurate timing.

Data Analysis:

Comprehension Data Analysis. Participants with <70% accuracy across all experimental items were excluded from analyses, this applied to 2 participants. Comprehension question accuracy was analysed with a mixed effects model in R with a binomial distribution. *Embedding* and *Extraction* were Fixed Effects and *Item* and *Participant* were Random Effects. Both random intercepts and slopes were included.

Reading Time Analysis. Any words with Reading Times (RT) greater than 2500ms or less than 100ms were eliminated from the data analysis. Additionally two items had typos in one word and those words were eliminated from analyses. Remaining RTs were log transformed and residual reading times (RRT) were calculated based on the difference between a word’s logRT and its predicted logRT determined through a model that included word character length (considering both experimental and filler items) as a Fixed Effect and Participant Random Effects (intercept only). This is a standard procedure for self-paced reading data and works to eliminate the confounding effects of

individual reading speed and character length per word (Ferreira & Clifton, 1986; Hofmeister, 2011).

There were five regions of interest for analysis: (1) the relative clause, excluding the preposition phrase, (2) the embedded verb (including any auxiliaries and/or prepositions) (3) the matrix verb (including any auxiliaries and/or prepositions), and the (4) head of the relative clause, for comparison with our production experiments (2A and 2B). In production, we expect to replicate Scontras et al.'s (2014) result that object filler-gaps result in longer production durations of the filler (ie, head of the relative clause, the reporter) than subject filler-gaps, whereas in self-paced reading we do not expect to have any effects, given that at that point there is no difference between an object and subject relative. Zones that included words that were eliminated because they were longer than 2500ms or less than 100ms in duration were excluded from all analyses. This resulted in 0.4 – 4% percent of the data being eliminated, depending on the specific region. We also considered zone outliers defined according to participant, condition, and region whereby the residual logRT was greater than or less than the mean $\pm 2.5SD$ (Tang, K. (2014). Linger Toolkit. <http://tang-kevin.github.io/Tools.html>). However, none were identified. Residual logRTs were analysed with a linear mixed effects model in R with *Embedding* and *Extraction* as Fixed Effects and *Item* and *Participant* as Random Effects. Both random intercepts and slopes were included. The p-values were generated by stargazer (Hlavac, 2014; stargazer: LaTeX/HTML code and ASCII text for well-formatted regression and summary statistics tables. R package version 5.1. <http://CRAN.R-project.org/package=stargazer>).

Results:

There were no significant effects in terms of comprehension accuracy (see Figure 1).

In terms of reading time, we found just like Gibson et al. (2005) that reading times of the relative clause were longer in object-extraction than subject extraction, although the significance was only marginal (see Figure 2^v; $B = -0.064$, $SE = 0.03$, $t = -1.91$, $p < .1$)^{vi}. We replicated the effect that embedded verb had a longer reading time in object extracted relative clauses than subject extracted ones (see Figure 2; $B = -0.069$, $SE = 0.013$, $t = -5.45$, $p < .01$)^{vii}. We also replicated the effect interpreted by Gibson et al. as being due to information flow: We found the relative clause to be read significantly slower in right-branching than center-embedding (see Figure 2; $B = 0.090$, $SE = 0.030$, $t = 3.04$, $p < .01$). The effect of attachment (center-embedding vs. right-branching) on reading times at the embedded verb was not significant, although reading times of the verb was numerically longer in right-branching than center-embedding. There was no significant interaction between embedding and extraction on reading times at the embedded verb. Given the non-significant effect of embedding at the embedded verb, but significance of the relative clause reading time, we investigated whether the relative-clause-effect was driven by an early effect, at the relative pronoun. Indeed, the embedding effect on reading times at the relative pronoun was highly significant (Figure 2; $B = 0.062$, $SE = 0.018$, $t = 3.41$, $p < 0.01$) with longer readings in right-branching than center-embedding. No other effects on reading times at the relative pronoun were significant. Another novel result was that the main verb had a significantly longer reading time in center-embedding than right-branching (see Figure 2; $B = -0.058$, $SE = 0.013$, $t = -4.47$, $p < .01$)^{viii}, neither the effect of extraction nor their interaction on reading times was significant at this region. Also, the head noun, that the relative clause modifies (see Figure 2), took significantly longer to read in right-branching than center-embedding ($B = 0.065$, $SE = 0.026$, $t = 2.50$, $p < 0.05$), but there was no effect of extraction nor an interaction between them.

Discussion

Experiment 1 replicated the main effects of Gibson et al. (2005), including the effect interpreted by Gibson et al. as being due to information flow, showing a longer relative clause duration in right-branching compared to center-embedded structures. This effect was taken as evidence in Gibson et al. that contrary to prior claims in the literature, right-branching structures are more difficult to process than center-embedded structures. However, we also found an effect pointing to a greater complexity for center-embedding compared to right-branching. This effect, however, is not observed during the relative clause, rather it is observable at the main verb (Gibson et al. 2005 did not report reading effects at the matrix verb). This is consistent with the integration component of the DLT, or some versions of similarity-based interference. Given there was no interaction with gap position (object vs subject), the cue-based account of Van Dyke and McElree (2006) that claims interference is driven by the shared feature of subject-hood is not supported. Furthermore, while Gibson et al. merely report a general effect of the relative clause having a longer reading duration in right-branching structures, we localized where this main effect of embedding is observed. We found that this greater duration is observed at an early point in the relative clause (ie, at the relative pronoun and not the main verb). In order to better understand and interpret the nature of these effects, we will now look at a data from a Planned Production study using the same stimuli.

Experiment 2A

Methods

Participants:

25 native American English speakers (21.25 years old on average and 13 were female) participated in the study. All participants gave informed consent in accordance with McGill ethics and were compensated for their participation.

Material: The same experimental stimuli to those of Experiment I were used with the exception of the British English modifications.

Procedure:

In the Planned Production procedure the entire sentence is displayed in the center of the screen and participants are instructed to read the sentence as many times as they wish until they are prepared to produce it aloud. Following production of the sentence they provided an acceptability rating of the sentence on a 5-point scale (1 is completely unacceptable and 5 is completely acceptable). This offline judgment task provides a rough approximation of syntactic complexity.

Data Analysis:

The sound files were verified for accuracy of content, trimmed for silence at the beginning and end, and underwent a forced-Alignment using HTK, trained on several hours of lab speech (using the prosody.lab forced-aligner, (Gorman, Howell, & Wagner, 2011). The durational zones were the same as in Experiment 1 (matrix verb, relative clause (excluding PP), embedded verb, head of relative clause). The regions were analysed for outliers based on mean \pm 2.5SD per participant, condition, and zone (Tang, K. (2014). Linger Toolkit. <http://tang-kevin.github.io/Tools.html>). No outliers were identified. Log duration of each zone was analyzed with a mixed effects model in R where *Participants* and *Items* were treated as random effects and *Extraction*, *Embedding*, and *phonemelength (scaled)* were treated as Fixed Effects.

Proportion of pauses prior to these zones were analysed with mixed effects logistic regression with *Extraction* and *Embedding* as Fixed Effects and *Participant* and

Items as Random Effects. Additionally, both duration of silence (excluding non-silence values, ie 0) and the acceptability ratings were analysed with a mixed effects model in R where *Participants* and *Items* were treated as random effects and *Extraction*, and *Embedding*, were treated as Fixed Effects. Both random slopes and intercepts were included unless convergence could not be met. In the discussion of the results we identify where the model was simplified on account of convergence failure. The p-values were estimated based on Wald's z-test (tables were generated by *Stargazer*, Hlavac, 2014, stargazer: LaTeX/HTML code and ASCII text for well-formatted regression and summary statistics tables. R package version 5.1. <http://CRAN.R-project.org/package=stargazer>).

Results & Discussion:

The production results replicate the self-paced reading finding from Experiment 1, as well as Gibson et al. (2005), the relative clause had a significantly longer duration with an object gap than a subject one (see Figure 3; $B = -0.03$, $SE = 0.01$, $t = -3.54$, $p < .01$). Furthermore, as in Experiment 1, we localized this effect to the embedded verb (see Figure 3; $B = -0.07$, $SE = 0.02$, $t = -3.24$, $p < 0.01$). Additionally, there was a marginal embedding effect on the embedded verb (see Figure 3; $B = 0.02$, $SE = 0.01$, $t = 1.86$, $p < .1$), showing a trend for the embedded verb to have a longer duration in right-branching than center-embedding. The interaction between extraction and embedding at the verb was again not significant. In replication of Scontras et al. (2014), we found the head of the relative clause had a longer duration with object extraction than subject extraction (Figure 3; $B = -0.038$, $SE = 0.010$, $t = -4.01$, $p < .01$). No other effects at the head of the relative clause were significant.

The production results do not, however, show a longer duration of the relative clause in right-branching compared to center-embedding, in contrast to the effect

observed in self-paced reading. Instead, we found a significantly higher proportion of pauses after the head of the relative clause in right-branching (see Figure 4; $B = 0.62$, $SE=0.21$, $z=2.92$, $p<.01$). There was also an effect of extraction on the proportion of pauses following the head of the relative clause (see Figure 4; $B=-0.56$, $SE=0.21$, $z=-2.67$, $p<.01$), whereby there were more pauses for object relative clauses. The interaction effect was not significant.^{ix}

Additionally, unlike Experiment 1, we did not find a significantly longer duration of the main verb in center-embedding than right-branching (or any other effects on this region). Instead, we saw a significantly higher proportion of pauses *before* the matrix verb (see Figure 5; $B = -2.77$ $SE = 0.27$ $z = -10.40$ $p <.001$)^x. Similarly, the duration of silence before the matrix verb was also significantly longer in center-embedding than right branching ($B = -0.08$, $SE = 0.03$, $t = -3.30$, $p <.01$). No other effects on the proportion of breaks or duration of silence before the matrix verb were significant.

We also looked at the acceptability ratings for the sentences. There was an effect of filler-gap type in the acceptability ratings analysis (see Figure 6; $B= 0.36$, $SE=0.14$, $t = 2.57$, $p <.05$), such that object gaps were rated as less acceptable than subject-gaps, compatible with the hypothesis that they are harder to process. There was neither an effect of Embedding nor an interaction between Embedding and Extraction on the acceptability ratings. In other words, there was no evidence that either embedding structure was globally more difficult in the acceptability results. These results are consistent with other studies that found no difference in judgment ratings across the embedding manipulation (Gibson & Fedorenko, 2013; Gibson & Thomas, 1996).

Experiment 2B

In interpreting the results of Experiment 2A, we were concerned about a potential confound in the embedding contrast: the presence/absence of ambiguity of attachment of the prepositional phrase in the RC in the stimuli from Gibson et al. (2005). This is also the contrast where we find a difference across the two methodologies. As Gibson et al. (2005) note, there is an ambiguity of attachment with respect to the prepositional phrase in the right-branching condition where the PP (ie, *on Tuesday*) can modify the main verb (ie *ignored*) or the embedded verb (ie, *attacked*). This ambiguity is not present in the center-embedding condition where the PP can only attach to the embedded verb. Given the difference in location of the embedding effect observed across the self-paced reading versus planned production studies, we wanted to be sure that the ambiguity was not playing into the effects observed in production. In self-paced reading, the attachment ambiguity would not affect any of the regions investigated because the parser can only become aware of it at the end of the sentence, thus it does not seem able to explain the effect of embedding effect on the relative clause reading duration in comprehension. In production, it may have an effect during a region of interest, as the producer would already be aware of the preposition phrase and have to make a decision regarding its attachment. Presumably, this decision would be established early on, rather than, say, following the head of the relative, where we see an embedding effect. Nonetheless, to be sure we ran the same study as in Experiment 2A with the preposition phrase eliminated from the stimuli.

Methods

Participants:

24 native American English speakers (22 years old on average and 15 were female) participated in the study. All participants gave informed consent in accordance with McGill ethics and received compensation for their participation.

Material: The stimuli were identical to those of Experiment 2A with the exception of the PP being deleted from the relative clause.

Procedure:

Same as Experiment 2A.

Data Analysis:

Same as Experiment 2A

Results.

The results of Experiment 2B replicate the effects of extraction observed in Experiment 2A. Compared to subject extraction, object extraction demonstrated a significantly longer duration of the relative clause ($B = -0.050$, $SE = 0.013$, $t = -3.85$, $p < 0.01$) and embedded verb ($B = -0.201$, $SE = 0.023$, $t = -8.57$, $p < .01$). And, again, there was no significant effect of embedding, nor an interaction between embedding and extraction on either of these two regions. In fact, the direction of the difference of the means on the embedded verb depending on embedding was opposite to that in Experiment 2A, with center-embedding tending to have a longer embedded verb duration than in right-branching ($B = -0.022$, $SE = 0.019$, $t = -1.177$, $p > .1$). The results on the embedded verb across Experiments suggest then that there is no systematic effect of the embedding manipulation in this time region.

With respect to the effects on the head of the relative, there was a difference to Experiment 2A: While there was again a significantly longer duration of the head of the relative clause in right-branching than center-embedding ($B = -0.075$, $SE = 0.016$, $t = 4.63$, $p < 0.01$), we only found a marginal effect of extraction ($B = -0.020$, $SE = 0.012$, $t = -1.68$, $p < 0.1$). The interaction was again not significant.

Experiment 2B replicated the finding that right-branching demonstrated a significantly higher proportion of breaks after the head of the relative clause compared

to center-embedding ($B = 1.36$, $SE=0.300$, $z= 4.53$, $p<0.01$), as in Experiment 2A. No other effects were significant for the proportion of breaks following the head of the relative clause. Unlike Experiment 2A, there was a significant effect of Embedding on the duration of silence after the head of the relative clause ($B = 0.046$, $SE = 0.016$, $t= 2.869$, $p<0.01$), whereby there was more silence after the head of the relative clause in right branching. No other effects of silence following the head of the relative were significant.^{xi}

With respect to effects at the matrix verb, experiment 2B replicated the finding that center-embedding had a significantly higher proportion of breaks before the matrix verb compared to right-branching structures ($B = -2.03$ $SE = 0.29$ $z = -6.95$ $p < 0.01$), and a longer duration of silence before the matrix verb ($B = -0.074$, $SE = 0.019$, $t = -3.97$, $p < 0.01$). No other effects were significant considering proportion of pauses or duration of silence before the matrix verb.

The matrix verb also tended to have a longer duration in center-embedding compared to right-branching (similar to the effect observed in self-paced reading), but this effect did not reach significance ($B = -0.024$, $SE = 0.014$, $t = -1.74$, $p < 0.1$).

Again, there was an effect of extraction in the acceptability ratings analysis ($B= 0.239$, $SE=0.115$, $t = 2.1$, $p < 0.05$), such that object-extracted relative clauses were rated as less acceptable than subject-extracted relative clauses, suggesting that they were globally harder to process. And again, there was neither an effect of Embedding nor an interaction between Embedding and Extraction on the acceptability ratings, suggesting that there was no global difference in processing difficulty based on embedding.

Discussion

Experiment 2B replicated three main findings from Experiment 2A: i) A greater proportion and duration of pauses before the main verb in center-embedded structures; ii) A greater proportion of pauses between the head noun of the relative clause and the relative clause in right-branching structures; iii) a greater duration of the embedded verb for object filler-gap dependencies. This suggests first that these effects are reliable, and second that they were not due to the potential attachment ambiguity in Experiment 2A.

The absence of an effect of embedding position in the acceptability ratings in both experiments is potentially meaningful, given the presence of such an effect for the filler-gap dependency distance. Center-embedded and right-branching structures do not seem to be globally different in how hard it is to process them, while longer filler-gap dependencies do seem to be globally more difficult. A concern for Experiment 2A is that different factors might have negatively affected the acceptability of center-embedding and right-branching. The acceptability of center-embedded relative clauses might have been negatively affected by the non-local dependency between the subject (head) and the verb. On the other hand, acceptability of the right-branching relative clause might have been equally reduced due to the ambiguity of attachment of the PP. The effects of these two factors on acceptability could then result in no difference across the position of embedding contrast. The fact that we still see no difference in acceptability between center-embedding and right branching in Experiment 2B (i.e. even once we eliminated the PP and its attachment ambiguity for the right-branching condition), suggests that the non-local subject (head) -verb dependency does not affect processing complexity. It is possible that such processing difficulty is not observable in the acceptability rating data, but then we need to explain why a difference is observable for filler-gap dependency distance.^{xiii}

Before turning to interpretations of the differences and similarities between self-paced reading and production, we want to elaborate on a few differences between Experiment 2A and 2B.

Experiment 2B found no significant effect of embedding on the duration of the main verb, but the effect of embedding was closer to significance (0.1) than in Experiment 2A. It is worth mentioning that this trend could point to an antilocality effect. Adding the preposition to the relative clause (or any other constituent for that matter) makes it comparatively *more* likely that the main verb will occur next. Antilocality effects have been observed in German in self-paced reading (Konieczny, 2000). Konieczny (2000) first observed faster reading times on the main verb in German, the more adjectival phrases (AP) that were added to a relative clause modifying the object. German is a subject-object-verb (SOV) language, thus there is more distance between the head of the object and the verb in the case of additional adjective phrases (AP), yet reading times at the verb were found to be faster. An explanation for this effect is that there is less *surprisal* at the verb or put slightly differently, there is more syntactic information to enable the prediction of the verb with additional APs (Hale, 2001; Levy, 2008). Thus, under this perspective, production durations at the matrix verb should be shorter when the PP is present, in the case of center-embedding, due to lower surprisal of the verb. The presence/absence of the same PP within the relative clause, however, would obviously have no effect on the main verb in the right-branching condition. Despite the fact that the producer knows what they are producing, it is not surprising to find antilocality effects also in production. Predictability of a word, in fact, is a well-known factor influencing production duration, with more predictable words leading to shorter durations than less predictable ones (see e.g. (Lieberman, 1963) and (Turk & Shattuck-Hufnagel, 2014) for review).

Predictability might come into play because speakers cannot plan out a complete utterance beforehand, and hence have to rely on the timely activation of the structures while speaking.

Another difference between Experiment 2A and 2b relates to the effect of filler-gap distance and embedding position on the head of the relative. While Experiment 2a found the same effect as Scontras et al. (2014), that is, a longer duration of the head of the relative clause when there was an object gap in the relative clause, this effect was only marginal in Experiment 2b. Also, in Experiment 2b we found a longer duration of the head noun of the relative clause in right-branching than center-embedding, an effect not observed in Experiment 2a, but observed in self-paced reading in Experiment 1. In production, the embedding effect then consistently surfaces around the head noun of the relative clause, but its precise location slightly varied (either on the head or the pauses following it). We see no reason for why the presence or absence of the PP would effect production duration of the head of the relative, and it is possible that this region is simply variable with respect to its robustness (gap type) or where exactly the duration effect is observed (embedding).

Overall, Experiments 2A and 2B demonstrate that the main effects in the production data are remarkably robust, and not contingent on the attachment ambiguity in Experiment 2A.

General Discussion

Across one self-paced reading study and two planned production studies we show similarities in how syntactic dependencies affect temporal measures. Specifically, a longer filler-gap dependency demonstrates longer reading and production durations at the point of dependency resolution (ie, embedded verb). However, these methods also demonstrate differences in the precise localization of *some* effects. There are three

effects of this type. First, the object filler-gap dependency demonstrates a longer production duration at the head of the dependency, but, as would be expected, no such difference is observed in comprehension. Second, a longer subject-verb dependency demonstrates a longer reading duration of the main verb but in production we see a greater likelihood of a pause and its duration before the main verb. Third, right-branched relative clauses have a longer reading duration than center-embedded ones, as was observed by Gibson et al. (2005), but in production we see a longer pause following the head of the relative clause in object modification. Both the similarities and distinctions provide informative insight into the nature of the processes required to resolve these two dependencies and the source of difficulty in multiple center-embedding. Each of these results is discussed in turn.

Filler-gap Dependency Resolution:

The distance effects for the filler-gap dependency seem relatively straight-forward to interpret. In both methods, we found a longer duration for the relative clause containing an object gap (ie, longer distance) compared to a subject one at the point of dependency resolution (ie, the embedded verb) in both domains. At the embedded verb, storage and/or integration costs or similarity based interference during cue-based retrieval is highest for the non-local object gap, where locality can be defined according to new discourse referents or referentially similar constituents. This adds to a large body of literature that finds a processing advantage for subject filler-gap dependencies in comprehension across various methods and populations. Thus, the results are compatible with the idea that integration costs, or similarity-based interference costs, are similarly incurred across comprehension and production at the embedded verb.

Effects of filler-gap dependency distance at the head of the relative clause demonstrates differences between the two methods. In production, we replicate the

effect of Scontras et al. (2014): a longer duration of the head of the relative clause when the clause contains an object gap compared to a subject gap. This is significant in Experiment 2a and marginal in Experiment 2b. The lengthening of the head with object relatives may be due to planning a more complex structure, as suggested by Scontras et al. (2014). Another possibility is that the longer production duration at the head NP corresponds to better encoding processes, which would reduce the likelihood of interference at the point of retrieval. Recall that the producer will engage in self-monitoring to ensure the produced message is the intended one and doing so effectively requires dependency resolution. Additionally, this lengthening of the head of the relative clause could have benefits to the listener. It could either allow for better encoding by the listener or even perhaps provide a signal that a more difficult structure is coming, potentially biasing their expectation towards an object relative. In reading, we see no difference across the filler-gap distance manipulation at the head of the relative. This is of course as expected: at this point, there is no difference across the conditions, and no way for the reader to guess which type of gap is coming up.

Our replication of Scontras et al.'s (2014) production duration effect at the head of the relative (ie, filler) across the filler-gap distance manipulation is interesting from a methodological point of view: Scontras et al. (2014) elicited production from a picture context, whereas we provided the sentence and then had participants produce the sentence. Our results suggest that even when the sentence is provided to the participant, an effect of planning an immediately upcoming complex structure is observed in production. The marginal effect in Experiment 2b seems to suggest the robustness of the effect is variable, an observation, which requires further investigation.

MacDonald et al. (2016) interpreted Scontras et al.'s (2014) longer duration of the head of the relative clause in the object-relative condition to arise from competition

with an alternative structure (passive voice in the relative clause), equally produced in that context. This interpretation does not fare well for our results from Planned Production, where the participant is provided with a specific sentence to produce rather than *generating* a sentence that describes a picture context. Under the current conditions, competition should be eliminated (or minimal). Furthermore, the observed effect for this same contrast at the embedded verb and in the same direction is consistent with memory-based accounts. Overall, the data are compatible with Scontras et al.'s (2017) interpretation of increased memory demands in object compared to subject filler-gap dependencies and extends them on methodological grounds from the wh-question structure to relative clause ones.

Subject-Verb Dependency Resolution:

In the reading data, we observed an effect of modifier position, whereby the main verb was read for a longer duration in the center-embedding structure, in replication of earlier self-paced reading studies. This effect is often taken as evidence for an integration or similarity-based interference cost (Grodner & Gibson, 2005; Van Dyke & McElree, 2006).

In production, we find center-embedded structures have a greater proportion of pauses immediately *prior* to the verb. The production results, thus, seem to require an interpretation that is not based on integration/retrieval costs. It could be that the production result simply requires a *different* interpretation from the reading one, but it is also possible that the memory-based cost is actually not the right interpretation of the self-paced reading effect. This perspective receives support from our findings across the filler gap dependency manipulation. In that case, the greater integration cost for object compared to subject gaps is located at the embedded verb in both production and comprehension. Hence, if the effect at the main verb for the subject-verb dependency,

in reading, were due to greater memory demands we would expect to observe the effect in the same location in production.

The idea that the reading time effect at the main verb is due to a greater distance between the subject and the verb (in terms of the thematic role dependency) in center-embedded compared to right-branched structures also seems unmotivated on theoretical grounds. The relative clause forms part of the subject that is being assigned a thematic role, and is crucial in resolving its referent. From a semantic point of view, it therefore does not make sense to say that the subject is more distant from the verb if it includes a relative clause: it is adjacent to the verb whether or not there is a relative clause. At best, there is an increase between the matrix verb and the syntactic head of the subject, which it *agrees* with (overtly or not). In other words, center-embedding generates a greater linear distance between the (head of the) subject and verb, but this differs from the distance manipulation in the filler-gap dependency, where the distance is structural (as well as linear). Structural distance refers to the hierarchical relations between constituents. In filler-gap dependencies, both the filler and the intervening NP is in a hierarchically prominent position relative to the gap, i.e. they both c-command it. In the subject-verb agreement dependency, the head of the subject is in a hierarchically prominent position (ie, c-command position) relative to the main verb, but the embedded clause is not, it simply linearly intervenes the dependency. The relative structural vs linear position of the intervening material may play an important role in the demands memory mechanisms incur in resolving these dependencies. Support for this distinction is provided by syntactic evidence showing that c-command licenses syntactic relations in the case of filler-gap and agreement dependencies, but linear position is irrelevant.

We consider two related accounts for the higher proportion of pauses before the main verb in the production of center-embedded relative clauses compared to right-branched ones. In fact, as will be made clear, the two accounts might simply reflect different levels of analysis (semantic vs prosodic) of the same underlying source.

Prosodic junctures tend to follow a clause-boundary (pointed out by Adrian Staub p.c.). Thus the higher proportion of pauses prior to the main verb in center-embedding may reflect a prosodic break following the relative clause. While in the literature on the syntax-prosody interface (see Truckenbrodt (2007) for a review) such boundaries are typically simple reflexes of syntactic structure, e.g., reaching the end of a clause, prior psycholinguistic work has directly related them to a processing cost (Ferreira, 1991; Schafer, 1997). Schafer (1997) proposed the *Interpretive Domain Hypothesis*, which holds that “an intonational phrase boundary defines a point at which the processor performs any as yet outstanding semantic/pragmatic evaluation and integration of material.” Clause boundaries might correlate with such prosodic boundaries precisely because they are places that necessitate the wrap-up of outstanding semantic interpretation of the prior material. Thus, in center-embedded structures, the interpretation of the relative clause restriction of the subject head may manifest as a pause prior to the verb. The relative clause narrows down the referent from the broader set denoted by the head noun that it modifies (Altmann & Steedman, 1988; Crain & Steedman, 1985).

Watson & Gibson (2004), however, argued that the likelihood of a prosodic boundary between two words correlates simply with the *size* of the constituents, due to the greater processing cost that larger constituents induce. This model specifically predicts a high likelihood of a prosodic boundary at the end of a relative clause, simply

due to the fact that a constituent with a large number of words ends with the last word of the relative clause.

From the current data, it is not possible to disentangle to what extent the proportion of pauses we observe is due to the complexity of the relative clause restriction or more generally to the complexity of processing larger constituents once they are completed. A comparison with other types of clauses and modifiers would be necessary to disentangle the two interpretations, but to do so one would need to ensure that the complexity of the restriction does not vary with the size of the constituent, which is perhaps not a simple endeavour. Nonetheless, a previous self-paced reading study provides data that is informative to this distinction. Grodner and Gibson (2005) compared reading times at the verb across three conditions: no modification (*the nurse supervised*), relative clause modification (*the nurse who was from the clinic supervised*), and prepositional modification (*the nurse from the clinic supervised*). They found both modifications led to significantly longer reading times at the verb than no modification, but there was no significant difference between the two modifications, despite the relative clause containing more words than the prepositional phrase. Both the relative clause and prepositional phrase, however, make a restriction on the subject referent suggesting that the restriction is more relevant than the number of words prior to such durational effects. Using these same stimuli in a planned production task, we would predict that the likelihood/duration of a pause prior to the main verb would be equivalent across the two modifications, but greater than with no modification.

If this interpretation of the production result is correct, it also offers a new perspective on the reading result: In self-paced reading, the main verb is the point in time when a participant can be sure that the relative clause is over—and therefore this is

the point in time where she finishes computing the overall meaning of the subject, including the restriction of the relative clause, which would be observable as longer RTs in center-embedding in this modality. A memory-based account for the reading data, on the other hand, would have to evoke different explanations for self-paced reading and production. Existing results are at least compatible with this uniform novel hypothesis.

If the processing incurred by the relative clause is due to computing the restriction that manifests prosodically as a pause, we actually expect no global complexity difference between a subject-modifying and an object-modifying relative clause, as both involve a restriction (of an NP), but at different points in the sentence.^{xiii} Evidence compatible with this idea is provided by the acceptability ratings, from the current study and others, which fail to find a global processing difference between these two constructions, while the differences in filler-gap distance do lead to a significant difference in global acceptability. Our account would then predict a processing cost also at the end of the right-branched relative clause, which remains invisible in our studies since this constitutes the end of the sentence. A straightforward prediction is then that we should see evidence for this if we added additional material to the end of the sentence in the right-branching case, which cannot form part of the relative clause.

Independent of which of the above interpretations (clause, constituent size, or restriction) of the pause effect are correct, collectively the results are taken to demonstrate that resolving subject-verb dependencies is not taxing on memory mechanisms. While memory mechanisms can adequately capture the data arising from distance manipulations of a filler-gap dependency, they do not for those of the subject-verb type. The broader implication is that these two dependencies cannot be treated equivalently when it comes to memory mechanisms, contra current theories. The fundamental difference between these dependencies may come down to the fact that

object filler-gap dependencies allow for greater similarity between the “target” (ie, filler) and its “distractor” (ie, embedded NP) in terms of their structural position (ie, c-command) with respect to the point of dependency resolution (ie, gap/embedded verb). In subject-verb dependencies the “distractor” only linearly intervenes the dependency, but does not occupy a c-commanding relation with the main verb, as does the subject head. In line with syntactic-based theories (Grillo, 2008; Rizzi, 1990), it may be that structural intervention is significantly taxing on memory mechanisms, but linear intervention can easily be overcome. This may require that memory-based theories consider costs/interference not only of informational content but structural relationships as well.

Complexity of Center-Embedding vs Right-branching: What makes multiple Center-Embedding So Difficult?

The last section argued that computing the restriction of the relative clause is more costly than the subject-verb dependency. If this is so, then we still need to explain the difficulty with multiple center-embedding. Recall that there is greater processing complexity with multiple clauses that are center-embedded (5) than right-branched (6) (examples taken from (Gibson et al., 2005).

- (5) The student [who the professor [who the scientist collaborated with <the professor>] advised <the student>] copied the article.
- (6) The scientist collaborated with the professor [who <the professor> advised the student [who <the student> copied the article]].

While adding a second relative clause to the center-embedded one (5) causes the parser to essentially break down, it seems to have comparatively little impact in the case of right-branching (6). In considering our argument from the previous section, this breakdown would not seem attributable to a verb being more distant from its subject. Rather we would argue that the cost associated with multiple center-embeddings arises from there being two nested object relative clauses in (5) compared to two subject relative

clauses in (6). This points the problem away from the subject-verb dependency and towards the filler-gap dependency and particularly the greater difficulty with object gaps than subject gaps. Now consider (7) and (8).

(7) The man [₁that the woman [₂ that the dog bit <the woman>] likes <the man>] eats fish.

(8) I saw the man [₁ that the woman [₂ that the dog bit <the woman>] likes <the man>]

In (7) there are two center-embedded relative clauses, whereas in (8) there is a right-branched and center-embedded relative clause. Gibson & Fedorenko (2013) argue that (7) and (8) are comparably difficult to process. Both involve two nested restrictions, which each involve object-extracted relative clauses. This finding is consistent with the explanation that object filler-gap dependencies are the source of processing breakdown in multiple center-embedding.^{xiv} MacDonald (2013) also attributes the problem of multiple center-embeddings to object relative clauses rather than long distance subject-verb dependencies. The complexity of object relatives from her perspective is two-fold: (1) ambiguity with an alternative structure (subject relative in comprehension and passive in production) and (2) semantic interference between the two NPs. We agree with both of these points but would also extend interference to consider syntactic properties (ie, relative structural position).

Relative Clause Position and the Information Flow Account:

The two methods also give rise to a distinction in the location of an effect for subject vs object modification in the region of the relative clause. In self-paced reading, we found right-branching relative clauses have longer reading times than center-embedded ones, as did Gibson et al. (2005). However, in Planned Production we did not find such a durational effect, but rather a higher proportion of pauses following the

head of the relative clause in object modifying relative clauses than subject modifying relative clauses. The production result is consistent with data from a previous study (Watson & Gibson, 2004). An account in terms of information flow might account for the effect, but does not immediately explain why the effect should show up as a pause prior to the relative clause in production and longer reading duration through the relative clause in self-paced reading.

One possibility for the longer reading duration of the relative clause is that the reader might simply not expect a relative clause after the head noun of the object. In the absence of an expectation for a relative clause, the parser may want to close the clause (or verb phrase) as soon as possible. Then when the following word (ie, the relative pronoun) is revealed, the parser has to reopen the clause to embed the relative clause. The slow-down observed during the right-branching relative clauses may be for this revision process or formalized in terms of expectation-based (ie, surprisal) accounts the longer reading times would simply be due to the relative clause being less expected following an object than a subject. In line with this interpretation, is our finding that that the embedding effect appears early within the relative clause (ie, at the relative pronoun but not the embedded verb). This account is compatible with the ‘information flow’ account if the role of this hypothesis is simply to explain an expectation that there should not be a relative clause in object position. However, Gibson et al.’s (2005) original account suggests that the object modifying relative clause has a longer reading duration due to the cost of processing old information where new information is expected. This particular perspective would expect a longer reading duration throughout the relative clause and not just at the beginning, as we observed. It is also not clear how to define “old” vs “new” in Gibson et al. (2005), given the absence of a context, which makes the content of all relative clauses “new”. Nonetheless, a reason for the

dispreference for relative clauses modifying objects over subjects is needed. The Information Flow account may be on the right track in looking to information structure-based argument hierarchies. While both arguments of the main clause are definite, and in the out-of-the-blue context ‘new’, the argument with a relative clause modification could be argued to be more specific. Work has shown that there is preference for specific arguments to appear prior to non-specific (indefinite) ones, at least when non-canonical argument orders are involved (Titov, 2012). If this preference were made to apply to a scale of specificity rather than simply categorical classifications of specificity (definite vs indefinite), it could possibly account for the (dis)preference.

The idea that the longer relative clause duration with object modification is due to either an early-closure parsing strategy or a surprisal effect does not seem to explain a high likelihood of a pause following the head of the relative clause in right-branching. In right-branching (but not in center-embedding) the linear order is compatible with the relative clause being extraposed. We know independently from the placement of sentential adverbs that object-modifying relative clauses can be extraposed in principle (eg.: *John met the student yesterday that failed to submit his test*). The greater number of pauses in production might simply reflect string vacuous extraposition (cf. (Wagner, 2010) for such an account of prosodic boundaries separating the head from the following RC). Given that there is strong evidence that readers impose a silent prosody when reading (see Fodor, 2013 for a summary), the longer reading duration of the relative pronoun in object-modification may correspond to a break, marking the extraposition. Extraposition has been shown to be costly when not contextually expected in comprehension studies (Levy, Fedorenko, Breen, & Gibson, 2012), and has been shown to result in a greater likelihood of pauses and greater production duration of preceding material in Poschmann & Wagner (2016). Our data do not speak to whether

extraposition had an effect in our data, but it may provide an overarching account of both the production and comprehension data. Further investigation is needed whereby the presence/absence of extraposition is directly tested for.

Head of the Relative Clause:

We see some similarities across methods for the effect of embedding position. In the reading data, we see an effect of embedding position on the head of the relative, whereby its duration is longer in right-branching than in center-embedding. We observe a similar durational effect in production in Experiment 2b, but not 2a. It could be that this location corresponds to where a partial interpretation of the verb phrase is computed. This is the first point where the parser can choose the right idiomatic interpretation of a predicate, independent of the content of an upcoming relative clause (e.g., John threw the ball, John threw the party; in the spirit of these examples taken from Baker, Johnson, & Roberts, 1989) and literature cited therein). The head of the relative clause in object position might be a natural location for such semantic processing.

Conclusions

In conclusion, we found similarity between production and comprehension in terms of memory demands in the case of the filler-gap dependency, but not the subject-verb one. The results can be captured by either integration demands or similarity-based interference. We have also suggested that such accounts consider the relative structural position of the intervening material, as it may help in explaining the contrast in results between the two dependencies.

The effect of a longer subject-verb dependency (center-embedding relative clauses) surfaces as a longer reading duration of the main verb, and a pause before the main verb in production. The production results suggest that prosody may mark

completion of the relative clause restriction on the subject via a break. This explanation can also account for the effect on the main verb in self-paced reading, since this is the point in time at which the reader can be certain that the relative clause is completed. This novel hypothesis makes interesting predictions for the processing complexity of related structures, which should be pursued in future work. More broadly, the distinction across the two types of dependencies (filler-gap and subject-verb) requires their processing be treated distinctly. The current data show it is unclear that there is a greater recruitment of memory resources in the case of a longer subject-verb dependency.

The durational differences across embedding position in the relative clause region requires further investigation. We suggested that the effects observed in comprehension might be due to revision processes or *surprisal* during the relative clause where following an object (head) a simpler structure (eg, coordinated clause or subordinated clause) could be expected (measured as a longer duration of the relative clause). The effects observed in production might be a reflex of extraposition of the relative clause in object position to avoid embedding (indicated by the greater break after the head of the relative). Both effects might ultimately be due to a dis-preference for an embedded relative clause in object position, with the result being an extraposed structure.

At a methodological level, we take our results to be an illustration that the two paradigms, self-paced reading and planned production, can be fruitfully combined to study complexity effects. Just as eye-tracking while reading can lead to additional insights compared to self-paced reading since it provides multiple dimensions of analysis (e.g., regressive saccades vs. reading time, (Staub, 2010), production offers a

richer array of dependent measures, and the differences between the tasks can lead to novel mutually informative results.

Acknowledgements

We would like to thank Yosef Grodzinsky for critical discussions during the development of this project and acknowledge that support for this project was provided by a Social Sciences and Humanities Research Council (SSHRC) Standard Research Grant 410-2011-1062, *Relative prosodic boundary strength and its role in encoding syntactic structure* (MW), a Fundação para a Ciência e a Tecnologia (Portugal National Research Council) Research Grant PTDC/CLE-LIN/114212/2009, *Syntactic and lexical factors in processing complexity* (NG), and funds from a Canada Research Chair in Speech and Language Processing (MW). An earlier version of this work (Experiment 2A) was presented at ETAP2 (2nd Experimental and Theoretical Advances in Prosody) held at McGill University in 2011, we thank the audience for useful comments and suggestions.

References

- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30(3), 191-281. doi:10.1016/0010-0277(88)90020-0
- Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, 54, 218-250. doi:10.1016/j.cogpsych.2006.07.001
- Baird, R., & Koslick, J. D. (1974). Recall of grammatical relations within clause-containing sentences. *Journal of Psycholinguistic Research*, 3, 165-171. doi:10.1007/BF01067574
- Baker, M., Johnson, K., & Roberts, I. (1989). Passive arguments raised. *Linguistic Inquiry*, 20(2), 219-221.
- Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language*, 6, 183-200. doi:10.1017/S0305000900002245
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355-387. doi:10.1016/0010-0285(86)90004-6
- Bock, J. K., Dell, G. S., Garnsey, S. M., Kramer, A. F., & Kubose, T. T. (2007). Car talk, car listen. In A. S. Meyer, L. R. Wheeldon, & A. Krott (Eds.), *Automaticity and control in language processing* (pp. 21-42). Hove, England: Psychology Press.
- Chomsky, N., & Miller, G. A. (Eds.). (1963). *Introduction to the formal analysis of natural languages* (Vol. 2). New York, NY: Wiley.
- Clark, H. H., & Malt, B. C. (1984). Psychological constraints on language: a commentary on Bresnan and Kaplan and on Givon. In W. Kintsch, J. R. Miller, & P. G. Polson (Eds.), *Method and tactics in cognitive science* (pp. 191-214). Hillside, NJ: Erlbaum.
- 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. Karttunen, & A. Zwicky (Eds.), *Natural Language Parsing* (pp. 320-358): Cambridge University Press.
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30, 210-233. doi:10.1016/0749-596X(91)90004-4
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368. doi:10.1016/0749-596X(86)90006-9
- Fiebach, C. J., Schleuisky, M., Lohmann, G., von Cramon, D. Y., & Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. *Human Brain Mapping*, 24(2), 79-91. doi:10.1002/hbm.20070
- Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27(2), 285-319. doi:10.1023/A:1023258301588
- Fodor, J. D. (2002). *Prosodic disambiguation in silent reading*. Paper presented at the Proceedings of the North East Linguistic Society.
- Fodor, J. D. (2013). Pronouncing and comprehending center-embedded sentences. In M. Sanz, I. Laka, & M. K. Tanenhaus (Eds.), *Language Down the Garden Path*. Oxford: Oxford University Press.
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, 5, 519-560. doi:10.1007/BF00138988
- Friedmann, N. (2008). Traceless relatives: Agrammatic comprehension of relative clauses with resumptive pronouns. *Journal of Neurolinguistics*, 21(2), 138-149. doi:10.1016/j.jneuroling.2006.10.005
- Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, 119, 67-88. doi:10.1016/j.lingua.2008.09.002
- Garaffa, M., & Grillo, N. (2008). Canonicality effects as grammatical phenomena. *Journal of Neurolinguistics*, 21, 177-197. doi:10.1016/j.jneuroling.2007.09.001

- Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, 8, 8-11. doi:10.1016/j.tics.2003.10.016
- Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, 17, 684-691. doi:10.1111/j.1467-9280.2006.01767.x
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1-76. doi:10.1016/S0010-0277(98)00034-1
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & 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(2), 313-354. doi:10.1515/cogl.2005.16.2.313
- Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28, 88-124. doi:10.1080/01690965.2010.515080
- Gibson, E., & Thomas, J. (1996). *The processing complexity of English center-embedded and self-embedded structures*. Paper presented at the Proceedings of the NELS 26 Workshop on Language Processing, MIT Working Papers in Linguistics., Cambridge, MA.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(6), 1-13. doi:10.1037/0278-7393.27.6.1411
- Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-Aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics*, 39(3), 192-193.
- Grillo, N. (2008). *Generalized Minimality: Syntactic Underspecification in Broca's Aphasia*. (Doctoral dissertation), Utrecht University, Utrecht.
- Grillo, N. (2009). Generalized minimality: Feature impoverishment and comprehension deficits in agrammatism. *Lingua*, 119, 1426-1443. doi:10.1016/j.lingua.2008.04.003
- Grodner, D. J., & Gibson, E. A. F. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29, 261-291. doi:10.1207/s15516709cog0000_7
- Grodzinsky, Y. (2000). The neurology of syntax: language use without Broca's area. *Behavioral and Brain Sciences*, 23(1), 1-21. doi:10.1017/S0140525X00002399
- Hakes, D. R., Evans, J. S., & Brannon, L. L. (1976). Understanding sentences with relative clauses. *Memory and Cognition*, 4, 283-290. doi:10.3758/BF03213177
- Hale, J. (2001). *A probabilistic early parser as a psycholinguistic model*. Paper presented at the Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies.
- Hickok, G. (2010). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, 21, 1229-1243. doi:10.1162/jocn.2009.21189
- Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, 26, 376-405. doi:10.1080/01690965.2010.492642
- Holmes, V. M. (1973). Order of main and subordinate clauses in sentence *Journal of Verbal Learning and Verbal Behavior*, 12, 285-293. doi:10.1016/S0022-5371(73)80072-6
- Kempen, G., Olsthoorn, N., & Sprenger, S. (2012). Grammatical workspace sharing during language production and language comprehension: Evidence from grammatical multitasking. *Language and Cognitive Processes*, 27(3), 345-380. doi:10.1080/01690965.2010.544583
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29(6), 627-645. doi:10.1023/A:1026528912821

- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.
doi:10.1016/0010-0277(83)90026-4
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177.
doi:10.1016/j.cognition.2007.05.006
- Levy, R., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, 122, 12-36. doi:10.1016/j.cognition.2011.07.012
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 1-45. doi:10.1207/s15516709cog0000_25
- Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6, 172-187.
doi:10.1177/002383096300600306
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 1-16. doi:/10.3389/fpsyg.2013.00226
- MacDonald, M. C., Montag, J., & Gennari, S. P. (2016). Are there really syntactic complexity effects in sentence production: A reply to Scontras et al. (2015). *Cognitive Science*, 40, 513-518. doi:10.1111/cogs.12255
- Miller, G. A., & Chomsky, N. (1963). *Finitary models of language users*. New York: Wiley.
- Momma, S., & Phillips, C. (2018). The relationship between parsing and generation. *Annual Review of Linguistics*, 4, 233-254. doi:10.1146/annurev-linguistics-011817-045719
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Science*, 11, 105-110.
doi:10.1016/j.tics.2006.12.002
- Pickering, M. J., McLean, J. F., & Branigan, H. P. (2013). Persistent structural priming and frequency effects during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 890-897. doi:10.1037/a0029181
- Poschmann, C., & Wagner, M. (2016). Relative clause extraposition and prosody in German. *Natural Language & Linguistic Theory*, 34(3), 1021-1066. doi:10.1007/s11049-015-9314-8
- Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, 9, 119-137. doi:10.1037/1076-898X.9.2.119
- Rizzi, L. (1990). *Relativized Minimality*: MIT Press, Cambridge, MA.
- Santi, A., Grillo, N., Grodzinsky, Y., & Wagner, M. (2011). *Planned production and self-paced reading of relative clause attachment*. Paper presented at the Experimental and Theoretical Approaches in Prosody (ETAP) 2: Prosody in Context, Montreal, Canada.
- Schafer, A. (1997). *Prosodic parsing: The role of prosody in sentence comprehension*. (PhD), University of Massachusetts, Amherst, Ma.
- Scontras, G., Badecker, W., & Fedorenko, E. (2017). Syntactic complexity effects in sentence production: A reply to Macdonald, Montag, and Gennari (2016). *Cognitive Science*, 41, 2280-2287. doi:10.1111/cogs.12495
- Scontras, G., Badecker, W., Shank, L., Lim, E., & Fedorenko, E. (2014). Syntactic complexity effects in sentence production. *Cognitive Science*, 39, 559-583.
doi:10.1111/cogs.12495
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116, 71-86. doi:10.1016/j.cognition.2010.04.002
- Titov, A. (2012). *Information structure of argument order alternations*. (PhD), University College London.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74, 209-253. doi:10.1016/S0010-0277(99)00069-4
- Tooley, K. M., & Bock, K. (2014). On the parity of structural persistence in language production and language comprehension. *Cognition*, 132(2), 101-136.
doi:10.1016/j.cognition.2014.04.002

- Traxler, M. J., Morris, R., & Seely, R. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69-90. doi:10.1006/jmla.2001.2836
- Truckenbrodt, H. (2007). *The syntax-phonology interface*. Cambridge: Cambridge University Press.
- Turk, A., & Shattuck-Hufnagel. (2014). Timing in talking: what is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B Biological Sciences*, 369, 20130395. doi:10.1098/rstb.2013.0395
- Van Dyke, J. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 407-430. doi:10.1037/0278-7393.33.2.407
- Van Dyke, J., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55, 157-166. doi:10.1016/j.jml.2006.03.007
- Van Dyke, J., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, 65, 247-263. doi:10.1016/j.jml.2011.05.002
- Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology*, 67, 1274-1304. doi:10.1080/17470218.2013.858363
- Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory*, 28, 183-237. doi:10.1007/s11049-009-9086-0
- Wanner, E., & Maratsos, M. (1978). *Linguistic theory and psychological reality* (Vol. An ATN approach to comprehension). Cambridge, MA: MIT press.
- Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19(6), 713-755. doi:10.1080/01690960444000070

Condition	Example Sentence
Center Subject (Local)	The reporter [who _ attacked the senator on Tuesday] ignored the president
Right Object (NonLocal)	The president ignored the reporter [who the senator attacked _ on Tuesday]
Right Subject (Local)	The president ignored the reporter [who _ attacked the senator on Tuesday]
Center Object (Nonlocal)	The reporter [who the senator attacked _ on Tuesday] ignored the president

Table 1. Example sentences from each condition

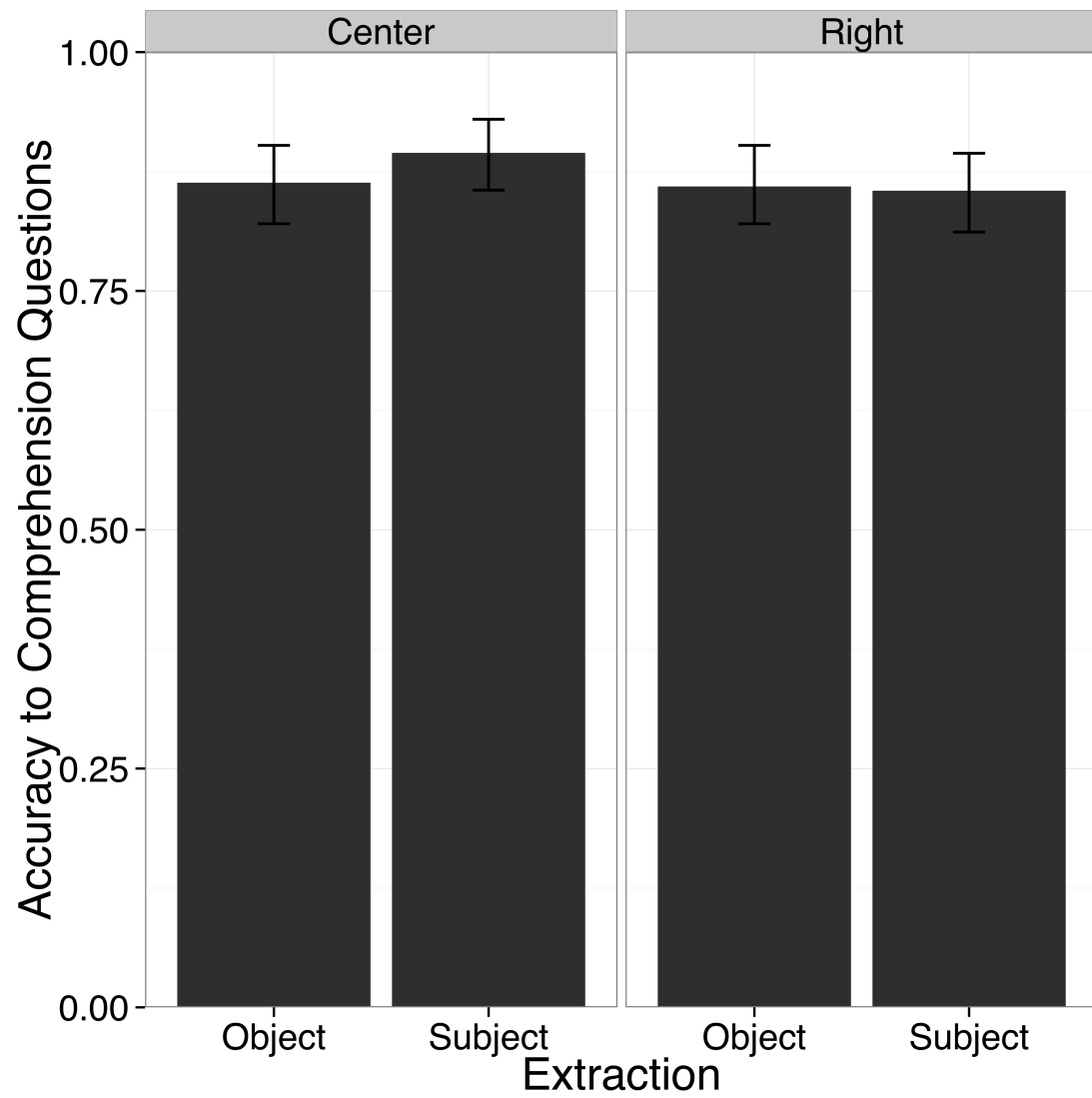


Figure 1.

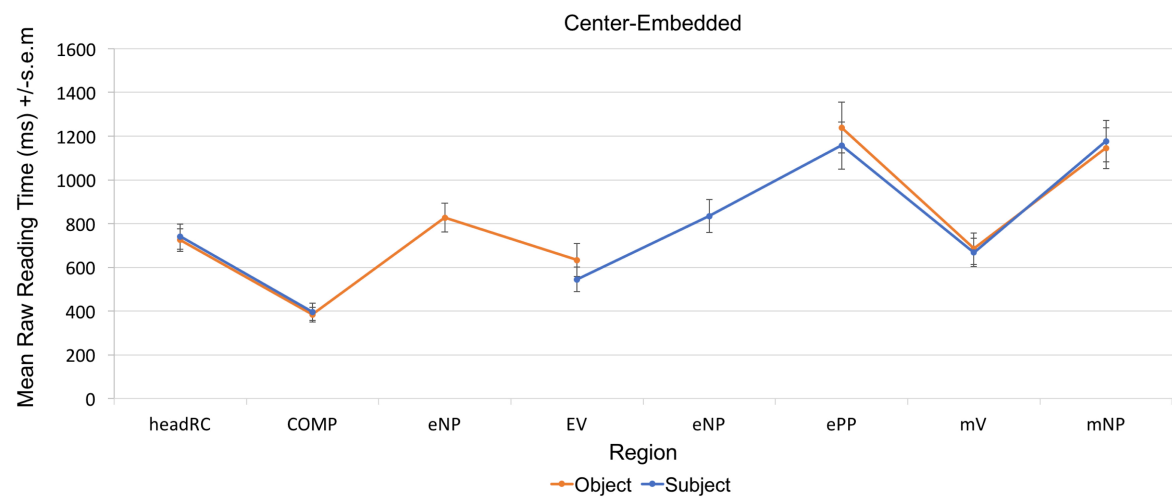


Figure 2A

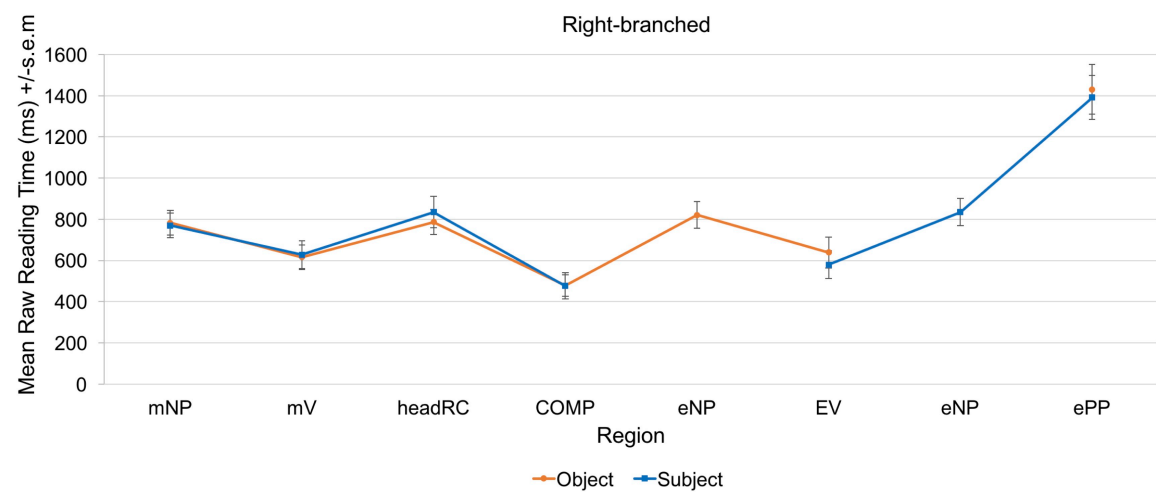


Figure 2B

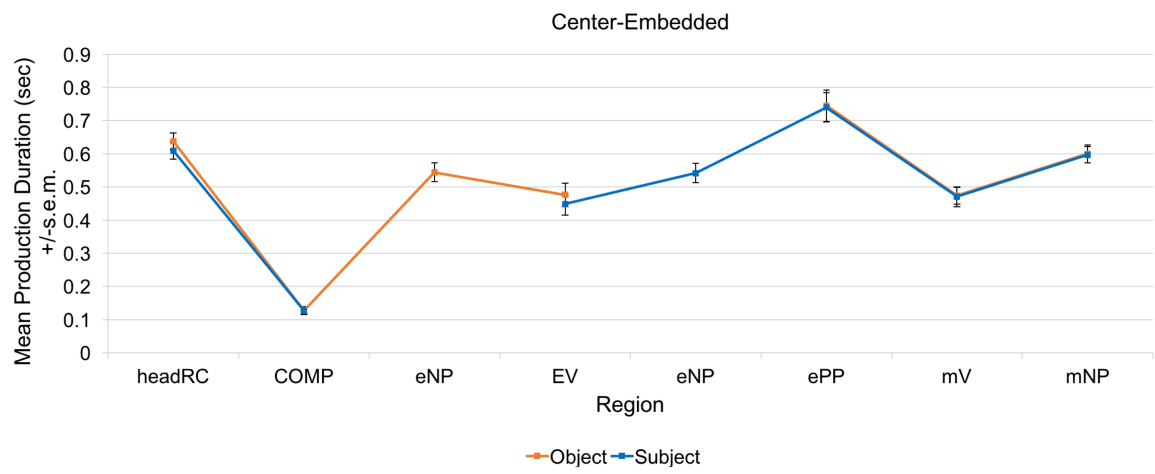


Figure 3A.

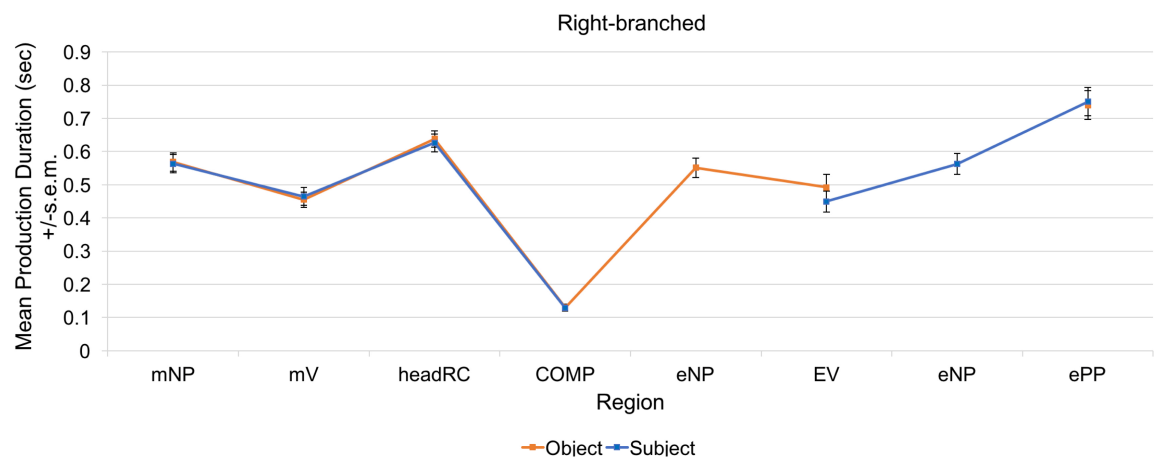


Figure 3B.

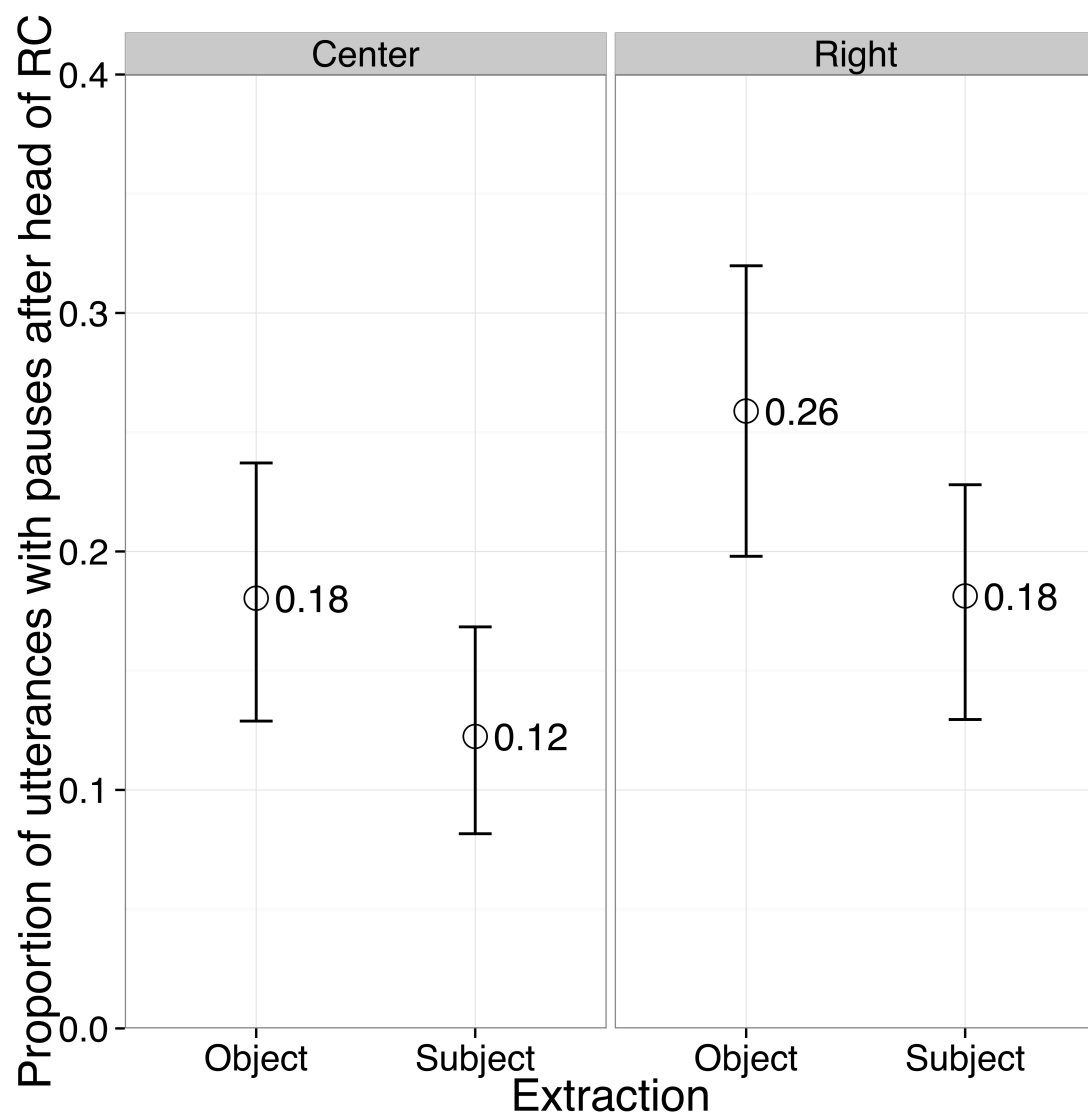


Figure 4.

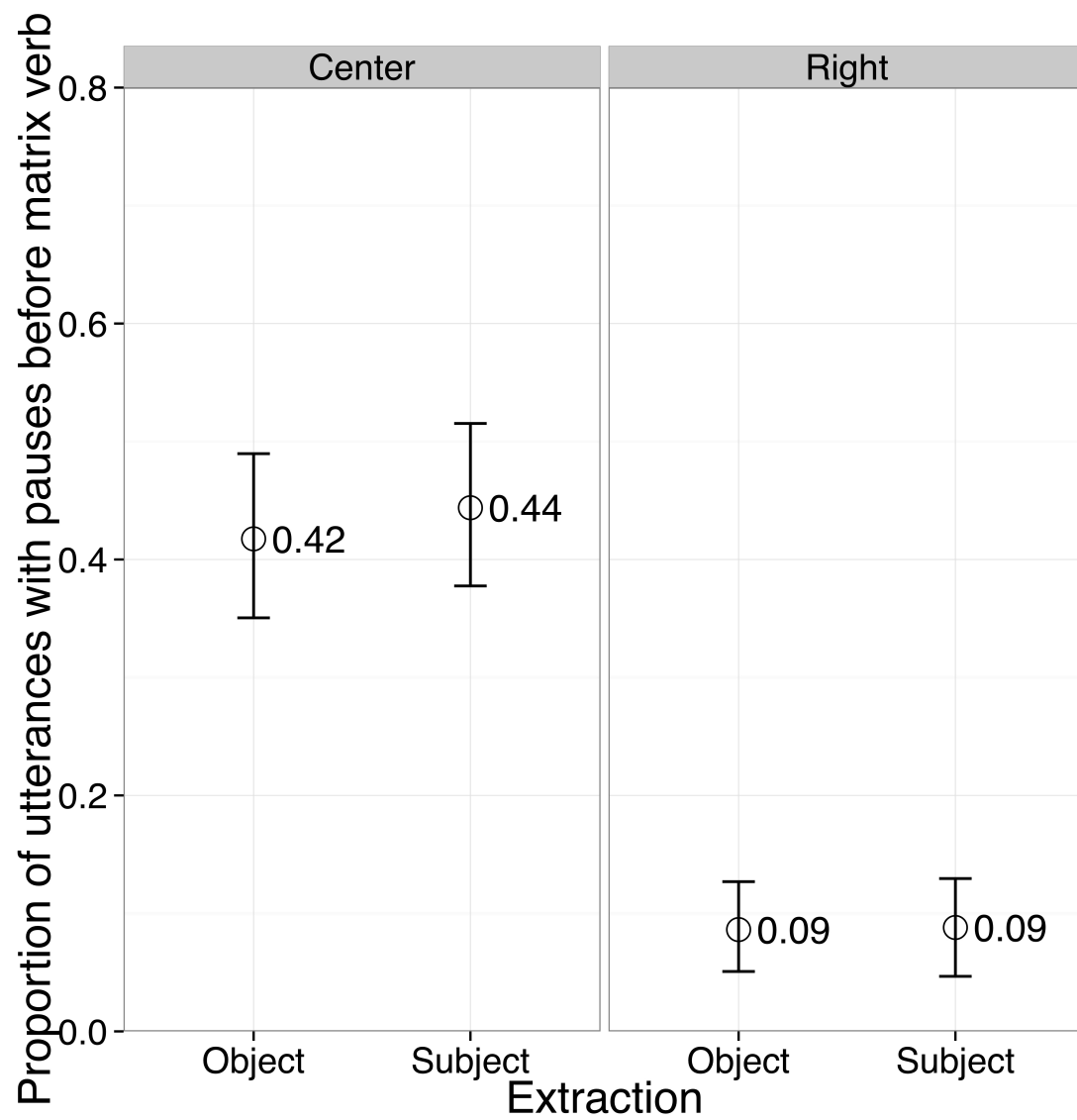


Figure 5.

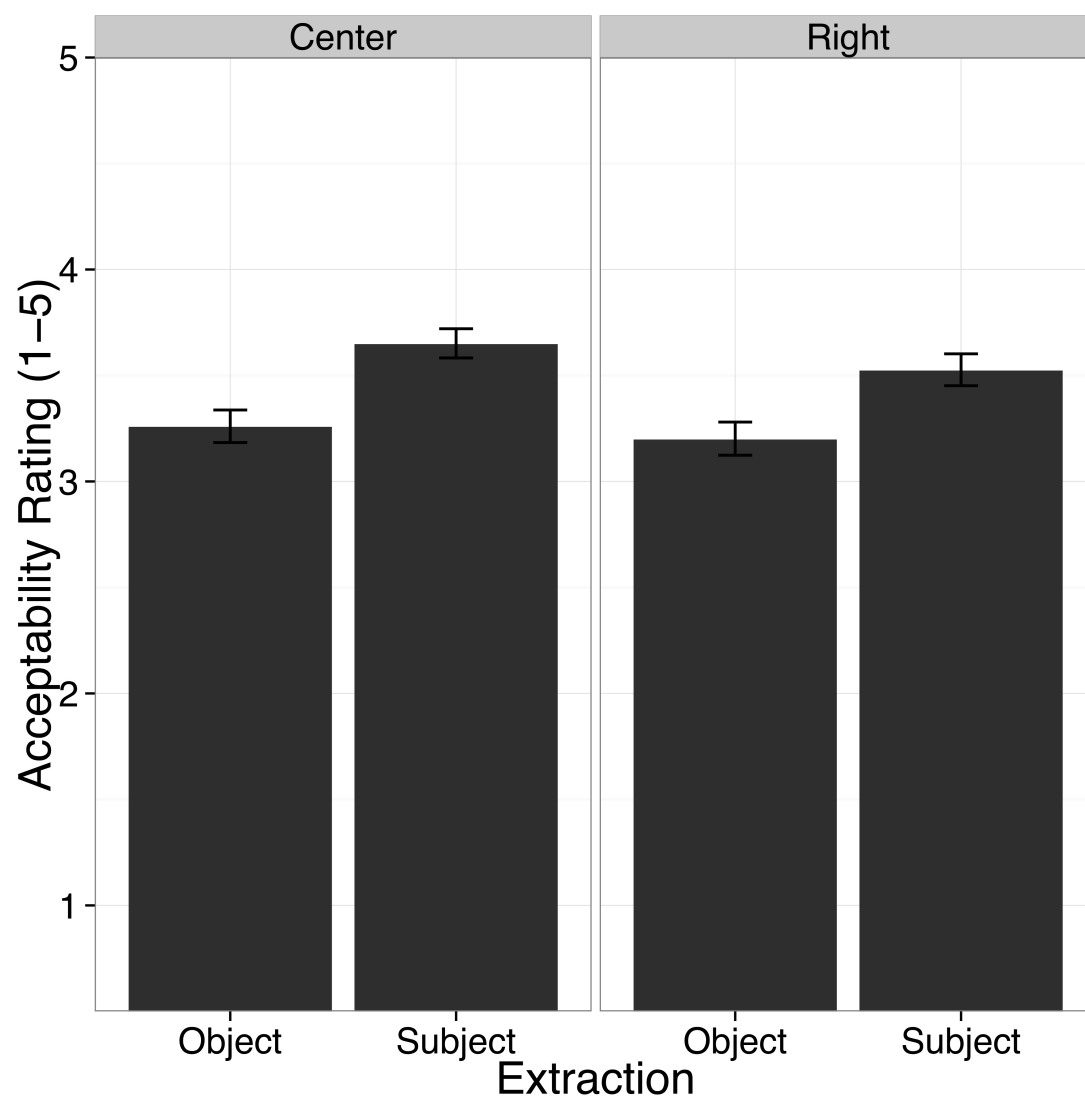


Figure 6.

Figure Captions

Figure 1. Mean proportion correct with confidence intervals to the comprehension questions in the self-paced reading study across conditions.

Figure 2. Mean raw reading times with standard error of the mean (s.e.m.) in center-embedded (2A.) and right-branched (2B.) structure across the sentence. Example sentence with region labelling: The reporter_{headRC} who_{COMP} (the senator_{eNP}) attacked_{eV} (the senator_{eNP}) on Tuesday_{ePP} ignored_{mV} the president_{mNP} (center-embedded)/The president_{mNP} ignored_{mV} the reporter_{headRC} who_{COMP} (the senator_{eNP}) attacked_{eV} (the senator_{eNP}) on Tuesday_{ePP} (right-branched)

Figure 3. Mean production duration with standard error of the mean (s.e.m.) in center-embedded (3A.) and right-branched (3B.) structure across the sentence. Example sentence with region labelling: The reporter_{headRC} who_{COMP} (the senator_{eNP}) attacked_{eV} (the senator_{eNP}) on Tuesday_{ePP} ignored_{mV} the president_{mNP} (center-embedded)/The president_{mNP} ignored_{mV} the reporter_{headRC} who_{COMP} (the senator_{eNP}) attacked_{eV} (the senator_{eNP}) on Tuesday_{ePP} (right-branched)

Figure 4. Mean proportion of pauses with confidence intervals following the head of the relative clause across conditions.

Figure 5. Mean proportion of pauses with confidence intervals prior to the matrix verb across conditions.

Figure 6. Mean acceptability ratings with confidence intervals in planned production (Experiment 2A) across conditions.

-
- ⁱ There is also some evidence that object relative clauses are more difficult because they are less expected than subject relatives, in accordance with expectation-based parsing. The evidence in support of these models, seems to be dependent on the paradigm, with evidence arising from eye-tracking Staub (2010). Here we use self-paced reading and focus on memory mechanisms, thus, we will not elaborate further on expectation-based parsing and long-distance dependencies despite it also playing a role.
- ⁱⁱ The relation between verb and the subject head is thematic. However, given the entire subject (not just the head) is assigned a thematic role, it is hard to understand the special status given to the head of the subject in order to provide a distance effect in this relation.
- ⁱⁱⁱ (Gibson & Fedorenko, 2013) argue a variation of (2) with the nested relative clause attached to the object is equally hard to process, which is unexpected if double-center-embedding is what creates the problem. We will return to this observation in the main discussion.
- ^{iv} Experimental sentences appeared without the first word capitalized and a final period. As this is true of all experimental conditions, it should not affect results and if anything should lead participants to think that is the interest of the experiment.
- ^v Figures of the self-paced reading data present raw reading times (msec), as this is more intuitive to understand. The analysis of raw reading times produced the same pattern of results.
- ^{vi} Model would not converge with full random effect structure. Analysis is based on random effects without the interaction term.
- ^{vii} Model would not converge with full random effect structure. Analysis is based random effects with intercepts only.

-
- ^{viii} Model would not converge with full random effects structure. Analysis is based on random effects including intercept and extraction.
- ^{ix} Apart from the proportion of pauses, we also looked at the *duration* of pauses, but we found no effect of Embedding, Extraction or their interaction on duration of silence after the head of the relative clause.
- ^x The model would not converge with random slopes. This model was therefore based on random intercepts alone.
- ^{xi} Why the effect on duration of silence was not observed in Experiment 2A is not clear and may simply be a Type II error. It is hard to conceive an explanation in terms of the presence/absence of the preposition phrase.
- ^{xii} Under the information-flow account, the lack of such an effect could be due to the incorrect flow of information in object-modified relative clauses equally reducing their acceptability.
- ^{xiii} At this point we have not addressed the longer duration of the RC in right branching than center-embedding, but we will attend to this later on in the discussion.
- ^{xiv} These predictions clearly need to be tested further. In an acceptability rating study, not reported here (based on a 6-point Likert scale), we found no difference when comparing the ratings between structures like (7) and (8), but with the lexical semantics better controlled, replicating Gibson & Fedorenko (2013). Further, in object-modifying environments, an object-extracted relative that nested a subject-extracted relative was rated more difficult than a subject-extracted relative modified by an object-extracted relative (all else equal). This indicates a critical factor for the integration cost. We could have included these rating studies in the final manuscript, but thought that this might divert too much from the main thread of this paper with its focus on self-paced reading and production.