Journal of Memory and Language 96 (2017) 93–109

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Appositives and their aftermath: Interference depends on at-issue vs. not-at-issue status

Brian Dillon [a,*], Charles Clifton Jr. [b], Shayne Sloggett [a], Lyn Frazier [a]

[a] Department of Linguistics, University of Massachusetts, Amherst, MA 01003, USA
[b] Department of Psychology, University of Massachusetts, Amherst, MA 01003, USA

ARTICLE INFO

Article history:
Received 28 June 2016
Revision received 26 April 2017
Available online 22 May 2017

Keywords:
Sentence comprehension
Working memory
Appositive relative clauses
Filler-gap processing
At-issueness

ABSTRACT

Much research has explored the degree to which not-at-issue content is interpreted independently of at-issue content, or the main assertion of a sentence (AnderBois, Brasoveanu, & Henderson, 2011; Harris & Potts, 2009; Potts, 2005; Schlenker, 2010; Tonhauser, 2011; a.o.). Building on this work, psycholinguistic research has explored the hypothesis that not-at-issue content, such as appositive relative clauses, is treated distinctly from at-issue content in online processing (Dillon, Clifton, & Frazier, 2014; Syrett & Koev, 2015). In the present paper, we explore the way in which appositive relative clauses interact with their host sentences in the course of incremental sentence comprehension. In an offline acceptability judgment, we find that appositive relative clauses contribute significantly less processing difficulty when they intervene between a filler and its gap than do superficially similar restrictive relative clauses. Results from two eye-tracking-while-reading studies suggests that recently processed restrictive relative clauses interfere to a greater degree with processes of integrating the filler at its gap site than do appositive relative clauses. Our findings suggest that the degree of interference observed during sentence processing may depend on the discourse status of potentially interfering constituents. We propose that this arises because the syntactic form of not-at-issue content is rendered relatively unavailable once it has been processed.

Introduction

The intuition that memory and attentional processes constrain and shape language processing has guided research in psycholinguistics since at least Miller and Chomsky (1963). Interestingly for the psycholinguist, these processes interact with linguistic structure in non-trivial and interesting ways. Very short center-embedded sentences may verge on the uninterpretable (Frazier & Fodor, 1978; Gibson, 1991; Kimball, 1973; Lewis, 1996; Miller & Chomsky, 1963), while certain right-branching or left-branching structures may grow quite large before placing any apparent strain on memory (Frazier & Fodor, 1978; Gibson, 1991; Kimball, 1973; Lewis, 1996; Miller & Chomsky, 1963; Resnik, 1992).

This state of affairs provides an interesting theoretical puzzle. In response, researchers have developed explicit proposals about the processing factors that contribute to sentence complexity (e.g. the Syntactic Prediction Locality Theory; Gibson, 1998 or the Dependency Locality Theory, Gibson, 2000; Warren & Gibson, 2002), as well as the nature of the memory architecture that supports sentence comprehension (Lewis & Vasishth, 2005; McElree, 2006; Van Dyke & Lewis, 2003). A virtue of this body of work is that it has produced a wealth of in-depth research on the fine-grained, incremental processing of syntactic and semantic dependencies, and has yielded valuable insights into the cognitive mechanisms that support sentence comprehension.

In the present work we seek to contribute to this program by investigating the incremental processing of appositive relative clauses in comprehension. Appositive relative clauses, like other parentheticals or supplements, are interesting to consider because in certain respects they are independent of their host clauses, in ways that we make precise below. Building on prior work, we develop several hypotheses for how appositives might interact with their host sentences during incremental sentence comprehension. To test these hypotheses, we present one offline rating study and two eye-tracking-while-reading studies that investigate how not-at-issue appositive relative clauses and at-issue restrictive relative clauses interact with filler-gap processing during reading. Our results suggest that during initial processing, not-at-issue appositives and at-issue relative clauses interact with their host

* Corresponding author at: Department of Linguistics, N408 Integrative Learning Center, University of Massachusetts, 650 North Pleasant Street, Amherst, MA 01003, USA.
E-mail address: brian@linguist.umass.edu (B. Dillon).

http://dx.doi.org/10.1016/j.jml.2017.04.008
0749-596X/

sentences in a similar fashion. However, once they have been processed, the two types of relative clause diverge in the amount of interference they create for processing syntactic dependencies in their host sentences. The findings we report here suggest that the amount of interference comprehenders experience during syntactic processing may depend in part on the discourse status of possible interfering constituents.

The at-issue/not-at-issue divide

Dillon, Clifton, and Frazier (2014) posed the question of how the distinction between at-issue and not-at-issue content (Potts, 2005) is reflected in sentence processing. Following Potts' terminology (Potts, 2005, 2015), *at-issue* refers to the main proposition asserted in a given utterance, a notion that dates back to Stalnaker (1978). At-issue content is sometimes understood as an answer to a question under discussion (QUD; see the notion of 'proffered content' in Roberts, 2012; see also Amaral, Roberts, & Smith, 2007; Tonhauser, 2011). At-issue content is contrasted with not-at-issue content. Our concern is with the content conveyed by parentheticals, nominal and clausal appositives, and expressives. Following Potts (2005), we will refer to this class of structures as *not-at-issue* content, and treat them as a class distinct from other non-asserted content such as presuppositions and implicatures (for a recent overview, see Potts, 2015). Examples are given in (1):

1. a. I met John – he's a real jerk, that one – while walking into town. (*parenthetical*).
   b. I met John, who had just gotten a haircut, while walking into town. (*appositive relative clause*).
   c. I met John, the new cheesemonger, while walking into town. (*adnominal appositive*).

Potts (2005) hypothesized that the not-at-issue content in in these examples is interpreted separately from the at-issue assertion *I met John while walking into town*. On Potts' view, each utterance in (1) has at least two distinct dimensions of meaning, the at-issue and the not-at-issue dimension. This independence explained several key features of the not-at-issue content: that it could be denied independently of the at-issue assertion (Amaral et al., 2007; Potts, 2005; c.f. Syrett & Koev, 2015), that it does not readily interact with semantic operators in the at-issue content (Amaral et al., 2007; Potts, 2005; but cf. Harris & Potts, 2009; Schlenker, 2010), and that it is often interpreted as a comment on the at-issue comment, rather than addressing the QUD or otherwise forming a coherent discourse with the material that surrounds it.

Although there is broad agreement that not-at-issue content is in important ways distinct from its host clause, there remain many unresolved debates about the precise way in not-at-issue and at-issue content are related. Some models hold that not-at-issue content is directly entered into the common ground (AnderBois, Brasoveanu, & Henderson, 2015, 2011; see similar ideas in Jasinskaja, 2016), other treat it as a pragmatic distinction (Harris & Potts, 2009; Potts, 2015), or others still treat it as a purely semantic distinction (Potts, 2005; Schlenker, 2010). More recently, some of the empirical claims made above have come under scrutiny: for example, Syrett and Koev (2015) report important experimental work raising the possibility that not-at-issue content may have a more substantial impact on the perceived truth conditions of the host utterance than is generally thought.

Dillon et al. (2014) asked whether the interpretive independence between not-at-issue material and their host clauses had consequences for incremental sentence comprehension. They asked whether syntactically complex material inside not-at-issue material (the underlined adnominal appositive in (2b)) contributed as much intuitive complexity to a sentence as did superficially similar material inside at-issue restrictive relative clauses (2a). In examples like (2), Dillon and colleagues manipulated syntactic complexity by the addition of an object relative clause (*Amy visited on Third Avenue*).

2. a. That butcher who was in the busy shop (Amy visited on Third Avenue) bought his meat from local farmers.
   b. That butcher, the one who was in the busy shop (Amy visited on Third Avenue), bought his meat from local farmers.

Across three acceptability judgment experiments, it was observed that the acceptability penalty associated with increased syntactic complexity was greater for restrictive relative clauses (2a) than for adnominal appositives (2b). This observation held whether the critical structures modified the subject or the object (Experiments 3 and 4), and held true whether the head noun was introduced by the definite determiner (Experiment 4) or a demonstrative (Experiments 1 and 3). It also held true when filler-gap dependencies, rather than object relative clauses, contributed the additional syntactic complexity (Experiment 1). This effect did not seem to reflect raters simply disregarding the content of the appositive structures out of hand: grammatical agreement errors inside the not-at-issue appositive were detected as readily as identical errors in at-issue, restrictive relative clauses (Experiment 2).

Dillon et al. (2014) hypothesized that this behavior was rooted in the fact that the not-at-issue content contributes a 'quasi-independent' speech act from its host clause, and that this quasi-independence has perceptual consequences for online sentence processing (for arguments that appositives and parenthetical asides contribute their own speech act, see Arnold, 2007; Frazier, Dillon, & Clifton, 2015; Syrett & Koev, 2015 dub this property *illocutionary independence*). If the appositive content is a quasi-independent speech act, then one might expect the processor to treat the appositive as if it were a partially distinct sentence in the middle of its host sentence. On this view, the processing of the appositive should be independent of its host sentence more or less to the extent that processing one sentence is independent of a distinct sentence in a discourse. Broadly speaking, this view predicts less interference between the at-issue content and the content of the adnominal structures in (2b) than between the at-issue content and the restrictive relative clause in (2a). This is because (2b) effectively breaks down into two sentences with relatively little overlap, while the same is not true of (2a). In other words, in (2a) the additional syntactic complexity complicates a single, integrated representation of the sentence and in (2b) that complexity is distributed across two distinct representations of the sentence, the more prominent of which is a syntactically simple at-issue clause. In this way the at-issue/not-at-issue partition explains why additional complexity inside the adnominal appositive imposes less of a penalty on judgments of sentence complexity than does complexity inside of a restrictive relative clause.

Distinguishing at-issue/not-at-issue content in processing

The offline judgment data suggest that the syntactic representations of the at-issue and not-at-issue material are to some extent independent of one other. However, these data raise interesting theoretical questions about how this independence arises during the course of incremental sentence processing. We can envision several distinct possibilities. One possibility develops a line of

thinking that goes back to Grosz and Sidner's (1986) model of dialogues. In their model, discourse units and their semantic content are associated with representational units they dubbed *focus spaces*. These consist of all the semantic information associated with a given unit. Grosz and Sidner modeled the attentional state of a speaker as a set of these focus spaces, with the availability of any individual focus space determined by its position in a stack memory structure. Applied to incremental sentence processing, this model essentially holds that only the current discourse unit is highly activated for the comprehender. If we extend this model to include the availability of the syntactic and lexical material associated with the active discourse unit, then we expect that the syntactic processing of one discourse unit should be relatively free of interference from the contents of other discourse units no matter where they occur, and no matter what relationship the discourse units bear to each other. For example, this view would lead us to expect that the processing of an appositive relative clause should be relatively free of any interference from the syntactic and lexical content of its host clause, as the appositive relative clause constitutes a distinct discourse unit from its host clause in the relevant sense (Arnold, 2007; Frazier et al., 2015; Jasinskaja, 2016; Loock, 2007; Syrett & Koev, 2015). We will refer to this as the *total isolation* hypothesis, which hold that at-issue and not-at-issue content are distinct from each other at all points in processing due to their status as independent discourse units (i.e. speech acts).

However, we can envision another possibility. It could also be the case that during incremental syntactic processing of the not-at-issue content, the at-issue content may itself remain active while a not-at-issue appositive is processed. This is the view endorsed by discourse models that allow subordinate discourse units to be co-active with their embedding discourse units. Appositive relative clauses and other supplements often elaborate upon or explain the content of their host utterances (see Jasinskaja, 2016; Loock, 2007; Potts, 2005 and references therein), and thus are in sense often discourse subordinate to their host clauses (this may be mostly clearly the case for sentence medial appositive relative clauses: see Jasinskaja, 2016). When such a subordination relationship obtains, both the host discourse unit and its subordinate unit are available for the purposes of anaphora or other discourse attachment (Jasinskaja, 2016; see also the *Right Frontier Constraint* of Polanyi, 1988). There is some psycholinguistic evidence that the lexical and syntactic material associated with superordinate discourse units remains active during the processing of a subordinate unit: Redeker (2006) showed using cross-modal priming that the material associated with a superordinate discourse unit could prime targets presented during a parenthetical (subordinate) discourse unit in a naming task if there was a discourse marker that explicitly signaled the digression. Models such as these hold that the distinction between not-at-issue content and at-issue content arises once the listener has completed a subordinate discourse unit and has returned to its host unit; at this point, the subordinate unit is rendered less active or less readily available (Jasinskaja, 2016; see also AnderBois et al., 2011, 2015; Syrett & Koev, 2015).

Again, one might extend claims about the availability of discourse units to include claims about the availability of the syntactic and lexical material associated with those units. On this view, the syntactic material in the host clause would remain active throughout the processing of a discourse-subordinate unit, and thus could create interference in the processing of that subunit. However, once the comprehender has processed the subunit and returned to the main clause, the material associated with the discourse-subordinate clause would be rendered relatively unavailable. This predicts a different incremental processing profile than that suggested by the total isolation hypothesis. Instead, on this view the content of a not-at-issue unit should be relatively unavailable to interfere with subsequent processing of the host sentence only once a comprehender has processed that not-at-issue content. In other words, having processed the not-at-issue content, comprehenders no longer need to maintain a representation of its syntactic form in memory, and so may let it decay or otherwise become unavailable. We will refer to this as the *differential loss* hypothesis.

The two views we've sketched here, the total isolation hypothesis and the differential loss hypothesis, make different predictions about how, and when, the processing of not-at-issue structures should interact with their host sentences in comprehension. We now turn to an investigation of these two hypotheses using the processing of filler-gap dependencies as a window into how at-issue and not-at-issue structures interact during incremental syntactic comprehension.

Filler-gap processing

To better understand how and when the independence of not-at-issue and at-issue content arise in processing, we investigated the processing of appositive and restrictive relative clauses in sentences like (3):

3. a. The butcher asked **who$_i$** the lady who$_j$ __$_j$ bought Italian ham was cooking dinner for __$_i$.
   b. The butcher asked **who$_i$** the lady, who$_j$ __$_j$ bought Italian ham, was cooking dinner for __$_i$.

The critical feature of the examples in (3) is the filler-gap dependency inside the embedded clause. To successfully interpret the sentence, the parser must associate the first filler *who* with its gap site after *for* (subscripts are provided to indicate the association of fillers with gap positions). The process of relating the *wh*-filler to its gap requires the comprehender to hold the filler in memory while processing material that intervenes between the filler and the gap, and to later retrieve or reaccess that filler when the dependent gap site is reached. Existing theories of structural complexity posit this process to be difficult at various points, for different reasons. In (3), however, that difficulty is compounded by the fact that the long-distance filler-gap dependency spans relatively complex material, including a second filler-gap dependency that is internal to the relative clause (indicated in (3) by *who$_j$*). In (3a), the parser must maintain an open filler across a restrictive relative clause; in (3b), the filler must be maintained across an appositive relative clause. The key empirical question at present is when, and to what extent, appositive and restrictive relative clauses interfere with the establishment of this filler-gap dependency. We focus on two potential loci of difficulty: the point at which the embedded relative clause is introduced, and the point at which the filler is integrated with the gap.

At the point of introducing the embedded relative clause (*who bought* in (3)), a second filler-gap dependency is opened and immediately resolved, creating a limited sort of center embedding. Previous reading time studies have shown that similar configurations create processing difficulty. For example, Grodner, Gibson, and Tunstall (2002) provided evidence from self-paced reading that maintaining an open filler-gap dependency across an embedded relative clause caused slowdowns throughout the relative clause, beginning at the complementizer that introduced the relative clause. They attributed this effect to storage load incurred by having two open dependencies at the point in the sentence (see Lewis & Vasishth, 2005, for an analysis of this pattern in a cue-based retrieval model). In addition, Obata, Lewis, Epstein, Bartek, and Boland (2010) showed using self-paced reading that

the processing of the embedded verb in the relative clause is made more difficult by having an open filler-gap dependency in the root clause, a finding they attributed to proactive interference created by having two fillers with similar case features in memory (Obata et al., 2010). Therefore, one predicted locus of difficulty is the beginning of the relative clause region *who bought* in sentences like (3).

A second locus of difficulty in constructing filler-gap dependencies is the point at which the filler is integrated with the gap. The integration of a filler with its gap is reflected in increased difficulty at the gap site in both ERP and reading time measures (Fiebach, Schlesewsky, & Friederici, 2001, 2002; Frazier & Clifton, 1989; Gibson & Warren, 2004; Gordon, Hendrick, & Johnson, 2001; Kaan, Harris, Gibson, & Holcomb, 2000; McElree, Foraker, & Dyer, 2003; Phillips, Kazanina, & Abada, 2005; Staub, 2010; Van Dyke & McElree, 2006; Wagers, 2008; Wagers & Phillips, 2014; Warren & Gibson, 2005). Critically, this integration difficulty is related at least in part to difficulty in retrieving the filler from memory: Van Dyke and McElree (2006) showed that the presence of interfering representations in memory negatively impacts the process of filler integration, a penalty they attributed to interference that arises from competing representations in memory at the point when the filler is retrieved (see also Kush, Johns, & Van Dyke, 2015). Fedorenko, Woodbury, and Gibson (2013) showed using a simultaneous free recall task and a sentence reading task that multiple presentations of a filler word strengthened its memory encoding and reduced integration difficulty at the gap site. Lastly, numerous studies have shown that the distinctiveness of the filler in memory modulates reading times at the gap site (Gordon, Hendrick, & Johnson, 2004; Gordon, Hendrick, Johnson, & Lee, 2006; Gordon, Hendrick, & Levine, 2002; Gordon et al., 2001; Hofmeister, 2011; Hofmeister & Vasishth, 2014; Warren & Gibson, 2005). Thus, a large body of behavioral and electrophysiological evidence supports the claim that (i) the process of relating a filler to its gap can be slow and effortful at the gap site itself and (ii) that this process is impacted by the number and quality of other representations in memory. Given these observations, we expect that the presence of an additional *wh*-element in memory (i.e. the relative pronoun *who* that introduces the restrictive or appositive relative clause) would interfere with the process of integrating the filler at the main clause gap site, because it overlaps in (morpho-)syntactic features (i.e. the *+WH* feature) with the target filler (see Obata et al., 2010; Wagers, 2008).

Returning to our hypotheses, they may be distinguished by measuring processing difficulty at the embedded RC region and the gap site. The total isolation hypothesis claims that the syntactic material associated with the host sentence is made unavailable as soon as the appositive clause is encountered, as the latter forms a distinct discourse unit. This means that any open filler-gap dependency would be less available to interfere with the processing of the appositive relative clause at this point in processing. Thus on this view, we would expect that any processing difficulty observed in the relative clause (RC) region (e.g. *who bought*) should be diminished for appositive conditions (4c) relative to restrictive conditions (4a). On the other hand, the differential loss hypothesis holds that the filler-gap dependency in the host sentence should interfere to a comparable extent in both restrictive and appositive structures: in both configurations, the syntactic material inside the host clause is available to a similar extent, and so should create similar processing difficulty at the RC region.

Note that the two hypotheses make superficially similar predictions about the difficulty of integrating the filler with the matrix gap site. The total isolation hypothesis suggests that while the restrictive relative clause may interfere with filler-gap integration, the appositive relative clause is not available and thus will not interfere. The differential loss hypothesis also predicts that the integration of the filler at its gap site should be easier in the appositive conditions than in the restrictive conditions. This is because the syntactic material associated with the appositive relative clause will have become relatively unavailable once it has been processed, and so will create less interference with the process of integrating the filler at its gap than in the case of a restrictive relative clause.

In the present study, we investigated these hypotheses using a mixture of offline and online experimental techniques. First, we tested whether appositive relative clauses do interfere less with filler-gap dependencies than do restrictive relative clauses using an untimed, offline judgment task. We adopted the experimental conditions presented in (4), using comparable sentences without filler-gap dependencies (−**WH**) as baselines to control for any differences between restrictive and appositive structures.

4. a. [**RESTR, +WH**]: The butcher asked who the lady who bought Italian ham was cooking dinner for.
   b. [**RESTR, −WH**]: The butcher asked if the lady who bought Italian ham was cooking dinner for her guests.
   c. [**APPOS, +WH**]: The butcher asked who the lady, who bought Italian ham, was cooking dinner for.
   d. [**APPOS, −WH**]: The butcher asked if the lady, who bought Italian ham, was cooking dinner for her guests.

To preview our findings, Experiment 1 provides evidence that restrictive relative clauses do interfere with the processing of a filler-gap dependency more than lexically matched appositive relative clauses. However, since the difference between the total isolation and differential loss hypotheses critically turns on when these effects arise in incremental processing, Experiments 2 and 3 investigate these sentences using eye-tracking-while-reading as an online measure of processing to determine when, and to what degree, intervening relative clauses interfere with filler-gap processing.

Experiment 1: offline ratings

In Experiment 1 we used intuitive judgments of acceptability to assess whether appositive and restrictive clauses contribute an equal amount of processing difficulty when they intervene between a *wh*-filler and its gap position. Offline acceptability judgments are reliably sensitive to processing difficulty associated with a sentence and so may be leveraged to investigate questions of processability (Fanselow & Frisch, 2006; Gibson, 1998; Warren & Gibson, 2002; Hofmeister, Jaeger, Arnon, Sag, & Snider, 2013; Schütze, 1996; Sprouse, 2008; Sprouse, Fukuda, Ono, & Kluender, 2011). In previous work, we showed that offline acceptability judgments are sensitive to the distinction between restrictive relative clauses and appositive structures (Dillon et al., 2014).

To estimate the relative difficulty of identifying a filler-gap dependency, we compare the critical conditions (*+WH* conditions; 4a, c) to minimally different control sentences that lacked a filler-gap dependency (*−WH* conditions; 4b, d). The *−WH* conditions control for the possibility of any baseline differences between restrictive and appositive relative clauses. For example, the contextual licensing conditions on restrictive and appositive relative clauses differ, leading to the possibility that there may be differences in naturalness between the restrictive and appositive RCs (see Grodner, Gibson, & Watson, 2005, among many others). However, if the two types of relative clause differ with respect to how available they are in memory once they've been processed, then we expect an interaction of *wh*-movement and relative clause type. Specifically, we expect there to be a smaller penalty on naturalness

judgments for having a filler-gap dependency that spans an appositive relative clause than one that spans a restrictive relative clause.

Methods

Subjects

Thirty-nine speakers of American English were recruited using Amazon's Mechanical Turk (https://www.mturk.com). Completion of the survey took approximately 20 min. Participation was restricted to workers with IP addresses in the United States, and all 39 participants self-reported as native speakers of English in a demographic survey. Participants gave informed consent, and were paid $5USD for their participation in the experiment.

Materials

We measured the acceptability of the four sentence types shown in (4). Our materials constituted a fully crossed 2×2 design, with the factors filler-gap dependency (*+WH* versus *−WH*), and RC structure (*APPOS* versus *RESTR*). All conditions consisted of bi-clausal structures, where the embedded clause was an embedded question. *+WH* conditions contained an embedded wh-question, and *−WH* conditions had embedded polar questions headed by *if*. *+WH* questions always had a bare *who* filler that was related to a gap site in a sentence-final position. In *−WH* questions this gap was filled with a fully lexical NP, with the exception of one item where the sentence-final gap site was replaced with an adverb (#5). In our experimental materials, the embedded subject was always modified by a relative clause, which was either restrictive (*RESTR*) or appositive (*APPOS*). Across all items, the appositive and restrictive clauses were always subject-extracted relative clauses. The relative clause always intervened between the *wh*-filler and the gap in *+WH* conditions. Appositive relative clauses were offset by commas, providing implicit prosodic cues that signal an appositive relative clause; restrictive relative clauses were never marked with commas. In all conditions, the relative clause structures were always introduced by the relative pronoun *who*. The full set of experimental items may be found in Appendix B.

Procedure

The 20 critical experimental items were distributed into four Latin Square lists and combined with 72 filler sentences. Participants were randomly assigned to a list. Fillers contained items from unrelated experiments, and comprised a range of acceptable and unacceptable structures of comparable complexity. Unacceptable items included unlicensed negative polarity item dependencies, ungrammatical reflexive dependencies, and ungrammatical agreement morphology. For each participant, the order of presentation was randomized.

The questionnaire was administered over the Internet using the IbexFarm experimental software (http://spellout.net/ibexfarm). Participants were instructed to rate sentences on a 1–7 Likert scale according to what sounded like 'natural' English, with higher values indicating greater naturalness. Only the end-points of the scale were labeled: 1 was *very unnatural* and 7 was *very natural*. Participants were given four examples to introduce them to the task. Each trial consisted of a single sentence presented on the screen, and the seven numerical response options were listed below the sentence. Participants responded by choosing their desired rating using either the mouse or the number keys on the keyboard. To remind participants of the interpretation of the Likert scale, the left end of the scale was labeled *Very unnatural*, and the right end of the scale was labeled *Very natural*. Participants were allowed to take as long as they liked to judge the sentences.

Analysis

Raw response data were analyzed using linear mixed effects models with the fixed-effects factors *structure* and *wh*, as well as their interaction, using the *lme4* package in the R statistical computing environment (Bates, Maechler, Bolker, & Walker, 2015; R Core Team, 2015). Crossed random intercepts and random slopes for all fixed effect parameters were included for subject and item grouping factors (Barr, Levy, Scheepers, & Tily, 2013). Simple difference coding was used (*APPOS* = 0.5; *+WH* = 0.5). We adopt the convention that a t value of absolute value greater than 2 presents a significant effect (Gelman & Hill, 2007). Where appropriate, we used the *lmerTest* package to extract pairwise comparisons from the fitted model objects (Kuznetsova, Brockhoff, & Christensen, 2015). In the calculation of the 95% CIs, the Satterthwaite approximation to the degrees of freedom was adopted.

Results

The average rating for each critical condition is presented in Table 1, along with by-participant standard errors. The results of the linear mixed effects model over the rating data is presented in Table 2. Analysis revealed a significant interaction of *structure* and *wh*, such that the effect of *wh*-movement was smaller for appositive structures. In addition, we observed significant effects of both the *structure* and *wh* fixed effects factors: stimuli with appositive relative clauses were rated more natural than stimuli with restrictive relative clauses by 0.39 points on the rating scale, and stimuli with *wh*-movement inside the embedded clause were rated as less natural than those without by 1.15 points on the rating scale.

Planned pairwise comparisons confirmed that this interaction was driven by a larger +wh penalty for restrictive RCs ($\bar{x} = 1.4$, 95% CI: [0.9, 1.9]) than for appositive RCs ($\bar{x} = 0.9$, 95% CI: [0.5, 1.3]).

Discussion

In Experiment 1, the presence of a filler-gap dependency inside an embedded clause negatively impacted the perceived naturalness of that sentence. In addition, the size of this filler-gap-penalty was modulated by the nature of the material intervening between the filler and the gap. Specifically, there was a larger penalty when the filler-gap dependency spanned a restrictive relative clause than when it spanned an appositive relative clause. This finding replicates and extends a pattern observed by Dillon et al. (2014). In their Experiment 1, they observed a similar interaction of *wh*-movement and intervener structure, such that the ratings

Table 1
Mean acceptability ratings for each condition, as well as marginal means, in Experiment 1. Parentheses represent standard error by participants.

	−WH	+WH	*Means*
APPOS	5.8 (0.2)	4.9 (0.2)	5.3 (0.2)
RESTR	5.6 (0.1)	4.2 (0.2)	4.9 (0.1)
Means	5.7 (0.1)	4.5 (0.2)	

Table 2
Results of linear mixed effects modeling of acceptability ratings in Experiment 1. Bold cells indicate significant coefficients.

	$\hat{\beta}$	t
structure	**0.39 (0.16)**	**2.47**
wh	**1.15 (0.19)**	**6.09**
structure:*wh*	**−0.49 (0.19)**	**−2.59**

penalty for *wh*-movement in the matrix clause was lessened when the dependency spanned a nominal appositive headed by *the one* than when it spanned a superficially similar restrictive relative clause.

We interpret the pattern of acceptability judgments to indicate that the presence of a filler-gap dependency leads to processing difficulty, which translates to a reliable decrement in offline naturalness measures (see also Dillon et al., 2014; Fanselow & Frisch, 2006; Gibson, 1998; Warren & Gibson, 2002; Hofmeister et al., 2013; Sprouse, 2008; Sprouse et al., 2011; *a.o.* for similar observations). Moreover, the pattern of results observed in Experiment 1 confirms our predictions that the processing difficulty associated with completing a filler-gap dependency is greater when a restrictive relative clause intervenes between filler and gap than when an appositive relative clause does. As in Dillon et al. (2014), we interpret this as evidence that raters are in some sense processing the appositive structure independently of its host clause.

We also observed a main effect of relative clause structure, indicating that restrictive clauses were overall rated less natural than appositive relative clauses. However, this main effect must be interpreted with caution due to the interaction of *wh* and *structure*. Within the non-movement baselines, there was a very small (0.2 point) difference between restrictive and appositives, suggesting that if anything raters perceived the appositives to be slightly more natural than the restrictive relative clauses in an out of the blue context.

The specific nature of this filler-gap penalty is of critical interest for the present hypotheses. In order to distinguish the total isolation and differential loss hypotheses, it is necessary to obtain an online measure of processing difficulty and evaluate when in the course of incremental processing the critical interaction is observed. We now turn to Experiment 2, where we used eye-tracking-while-reading to evaluate this.

Experiment 2: eye-tracking while reading

In Experiment 2, we used eye-tracking-while-reading to evaluate incremental processing difficulty associated with the sentence structures we examined in Experiment 1. We repeat the critical stimuli in (5) below, annotated with regions of analysis for the eye-tracking while reading experiment. Underscores mark the position of the gap in the *+WH* conditions, though these were not presented to participants in the experiment.

5. a. [**RESTR, +WH**]/The butcher/asked who/the lady/who bought/Italian ham/was cooking/dinner for _./
   b. [**RESTR, −WH**]/The butcher/asked if/the lady/who bought/Italian ham/was cooking/dinner for her guests./
   c. [**APPOS, +WH**]/The butcher/asked who/the lady,/who bought/Italian ham,/was cooking/dinner for _ ./
   d. [**APPOS, −WH**]/The butcher/asked if/the lady,/who bought/Italian ham,/was cooking/dinner for her guests./

Experiment 1 established that the complexity of the filler-gap dependency is modulated by the type of material that intervenes between the filler and the gap. However, the total isolation and differential loss hypotheses are distinguished on the basis of when this modulation occurs. On the total isolation hypothesis, we expect that the structure of the relative clause should interact with the presence of a filler-gap dependency at the beginning of the relative clause (e.g. *who bought*). In particular, we expect very little cost for an open filler-gap dependency on this region when it is inside an appositive relative clause, because the open filler-gap dependency is not kept active while comprehenders process the not-at-issue appositive relative clause. In contrast, the differential loss hypothesis posits no incremental distinction between the appositive and the restrictive at this point in processing, and so predicts that at this region we should observe that the presence of an open-filler gap dependency interferes to a similar degree with appositive and restrictive relative clauses alike at this region.

The differential loss hypothesis predicts that appositive and restrictive structures should be distinguished only once the relative clause has been processed, at the gap site in the host clause. If the material associated with appositive is less available than the material associated with the restrictive relative clause, then it will be less available to interfere with the gap-integration process, thereby easing integration difficulty at the gap site when the filler-gap dependency spans an appositive relative clause. This would result in an interaction of *wh* movement and *structure* at the point when readers complete the filler-gap dependency, the sentence-final region (*dinner for*).

Methods

Subjects

59 native speakers of American English were recruited to participate in Experiment 2; of these, 8 participants' data was removed prior to analysis due to unacceptable levels of track loss, leaving a total of 51 participants included in the final analysis. Participants were recruited in undergraduate Linguistics classes at the University of Massachusetts, Amherst. Participants gave informed consent and were given course credit in exchange for participation. Completion of the experiment, including calibration of the eye-tracker, took approximately 45 min.

Materials

The 20 critical items from Experiment 1 were used in Experiment 2. To increase the number of observations per condition in eye-tracking, an additional four items were constructed. The additional four items are listed in Appendix B (Items #21–24). The 24 target items were combined with 50 unrelated fillers. All filler sentences were of comparable length and complexity to the target items. All fillers were grammatical, and contained a range of referential and structural ambiguities.

Procedure

Participants read sentences while their eye movements were recorded with a tower-mounted Eyelink 1000 eye tracker (SR Research). Participants had binocular vision while movements were measured, but only the movement of the right eye was recorded. While seated, participants' heads were immobilized using a chin rest and forehead restraint that was adjusted for comfort. Subjects were seated approximately 35 in. from the monitor, giving them approximately 4 characters per 1° of visual angle. The sentences were presented one at a time on the screen in a black fixed-width font (Monaco, 11 pt) on a white background. All stimuli fit on a single line of text.

Before the experiment, the experimenter calibrated the eye-tracker with a 13-point display, and the experimenter ensured that eye position errors were no greater than 0.5°. Subsequent to calibration, participants were given four practice trials. Participants initiated each trial by fixating on a marker at the beginning of the sentence. Once a fixation on the marker was recognized by the experimental software, the target sentence was displayed all at once. Participants ended the presentation of the test sentence by pressing a button on a response pad when they had finished reading the sentence. A comprehension question was presented immediately after the test sentence, and participants indicated their response on the same response pad. All comprehension questions were two-alternative forced choice questions. Questions

probed both the content of the relative clause and the embedded predicate.

After completion of the experiment, each participant was given a brief questionnaire to determine whether they reliably used commas to distinguish appositive relative clauses from restrictive relative clauses. In this post-test, participants presented with 20 sentences that completely lacked all punctuation and capitalization. They were asked to provide their preferred capitalization and punctuation for each sentence. 10 items had unambiguous restrictive relative clauses, and 10 had unambiguous appositive relative clauses. Unambiguous restrictive relative clauses either were introduced by 'that' (e.g. 'the man that was jumping up and down'), or were attached to quantified NPs (e.g. 'none of the kids who got a lollipop'). Unambiguous appositive relative clauses were those attached to proper names (e.g. 'little suzie who tried to run away').

Analysis

For purposes of analysis, the sentences were split into the critical regions as in (5). The first region always corresponded to the matrix subject, the second was always the embedding verb and the associated complementizer or *wh*-word, and the third was always the embedded subject. The fourth region always consisted of the relative pronoun and the following verb, and the fifth region was always the rest of the relative clause structure (most commonly a direct object NP). Lastly, the sixth region was the embedded verb and its auxiliaries, whereas the seventh and final region was the remainder of the sentence which most often consisted of a direct object NP and a prepositional phrase. The prepositional phrase inside the final region was the position of the gap site in *+WH* conditions.

We present two early measures of reading, and one late measure. For early measures, we present first pass time (the sum of all fixation durations in a region before it was exited either to the left or to the right), and go-past times (the sum of all fixations from entering a given region before exiting that region to the right). For late measures, we present total viewing times (the total time spent reading the sentence). We present raw RTs in milliseconds in the text, however all statistical analyses were performed on log-transformed RTs to ensure normally-distributed residuals; the choice of log-transformed RTs was determined by using the Box-Cox procedure to evaluate which transformation of the data would yield normally distributed residuals in a linear model (Box & Cox, 1964; see Vasishth & Nicenboim, 2016). For continuous RT measures, all statistical analysis was carried out using linear mixed effects models with the same structure as those used in Experiment 1. Any post hoc tests were carried out using *lmerTest*, as in Experiment 1. For analyses of the discrete regressions-out measure, we used logistic linear mixed effects models.

Prior to statistical analysis, blinks and other artifacts were manually removed, as were any individual fixations shorter than 80 ms or greater than 1000 ms. Trials with track loss or blinks in first-pass reading at the critical region (here, the verb following the relative clause) were excluded. Participants who lost more than 25% of their data in this way were excluded from further analysis, leading to the rejection of 8 participants' data. Of the remaining participants, removal of trials with track loss led to a loss of 5.7% of the data. Lastly, additional first-pass times of greater than 2000 ms were excluded, as were total times of greater than 4000 ms.

Results

Offline responses

The results of the offline post-test revealed that participants in this experiment consistently used commas to distinguish appositive relative clauses from restrictive relative clauses. Unambiguously restrictive relative clauses were given initial and final commas at extremely low rates (4.2% and 8.5%, respectively). In contrast, unambiguously appositive relative clauses were given initial and final commas at very high rates (78% and 86%, respectively). This pattern confirms that for our participant population, punctuation provides an effective cue to relative clause structure.

Performance on the comprehension questions indicated that participants understood the sentences at a high rate. Accuracy on the filler items was 87%. Accuracy on experimental items ranged from 87% to 91% across conditions. Logistic LME modeling of the responses yielded no significant differences between conditions.

Reading time measures

The mean total viewing times in Experiment 2 are presented in Table 3, and their distribution is presented in Fig. 1. Table 4 presents the results of the LME analysis on log-transformed total viewing times. Statistical analysis revealed significant effects of both *wh* and *structure*, as well as a significant interaction of the two factors. Inspection of the means and distributions indicates that this was due to significantly longer total viewing times in the *+WH*, *RESTR* condition than all other conditions.

Raw first-pass and go-past times for each region are presented in Fig. 2; tables of means may be found in Appendix A. The results of LME analyses on log-transformed first-pass and go-past times are presented in Tables 5 and 6, respectively. The results reveal two main patterns.

First, in all regions prior to the gap site, we observe only main effects of *wh* or *structure*. At the embedded subject *the lady(,)*, there were significantly slower reading times for *+WH* conditions in both first-pass and go-past times. At the beginning of the relative clause (*who bought*), there were main effects of both *structure* and *wh* on go-past times. This reflected longer go-past times in *+WH* conditions, and longer go-past times for restrictives than for appositives. There was also a main effect of *structure* on first-pass times. There was a trend towards an interaction observed in first-pass times on this region, but this trend was not significant, and it was not evident in go-past times. The main effect of *+WH* persisted into the final region of the relative clause.

However, at the gap site (the final region), our analysis revealed an interaction of *wh* and *structure* in go-past times at the final region (*dinner for (her guests)*), in addition to main effects of both *wh* and *structure* (note that there were substantial length differences between *−WH* and *+WH* conditions at this region). Inspection of the means suggests that the interaction was driven by a difference in the magnitude of the effect of *wh* for *APPOS* and *RESTR* conditions: while *APPOS* and *RESTR* did not differ substantially in *−WH* conditions, *RESTR*, *+WH* conditions were 559 ms slower than *APPOS*, *+WH* conditions at this region. Post-hoc pairwise comparisons confirmed that the effect of *WH* was significantly larger for *RESTR* conditions ($\bar{x} = 0$, 95% CI: $[-0.1, 0.1]$) than for *APPOS* conditions ($\bar{x} = -0.2$, 95% CI: $[-0.4, -0.1]$). We also observed a large main effect of *WH* at the final region, such that *−WH* conditions took

Table 3
Mean total viewing times for Experiment 2. Parentheses represent standard error by participants.

	APPOS, −WH	APPOS, +WH	RESTR, −WH	RESTR, +WH
Total viewing time	4535 (173)	4554 (187)	4602 (170)	5260 (224)

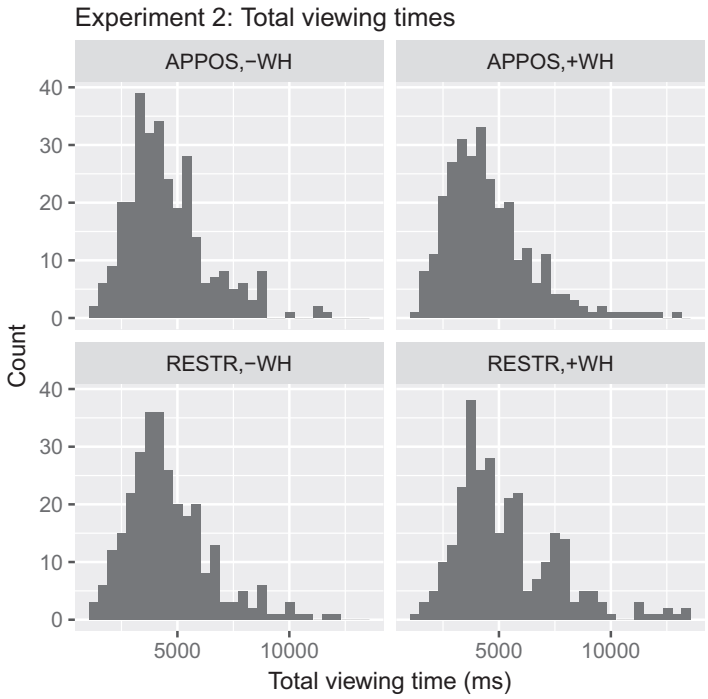

Fig. 1. Histogram showing the distribution of total viewing times by condition in Experiment 2.

Table 4
Summary of LME analyses of log total viewing times, Experiment 2. Bold cells indicate significant coefficients.

	TVT: $\hat{\beta}$	t
wh	**0.06 (0.02)**	**2.69**
structure	**−0.08 (0.02)**	**−3.18**
wh:*structure*	**−0.13 (0.04)**	**−3.52**

significantly longer to read. Presumably, this effect reflects the substantial length differences in this region between *−WH* and *+WH* conditions.

Discussion

We may summarize the results of Experiment 2 in the following manner. Early reading time measures indicated a slowdown at the subject immediately following a *wh*-pronoun (similar to a pattern reported in Staub, 2010, for object-extracted relative clauses). The difficulty associated with having an open filler-gap dependency persisted across the relative clause until the embedded verb. This difficulty did not interact with relative clause structure at any point prior to the gap. Turning to go-past times at the gap site, we observed an interaction of our experimental factors: *+WH* conditions were read significantly longer for *RESTR* conditions than for

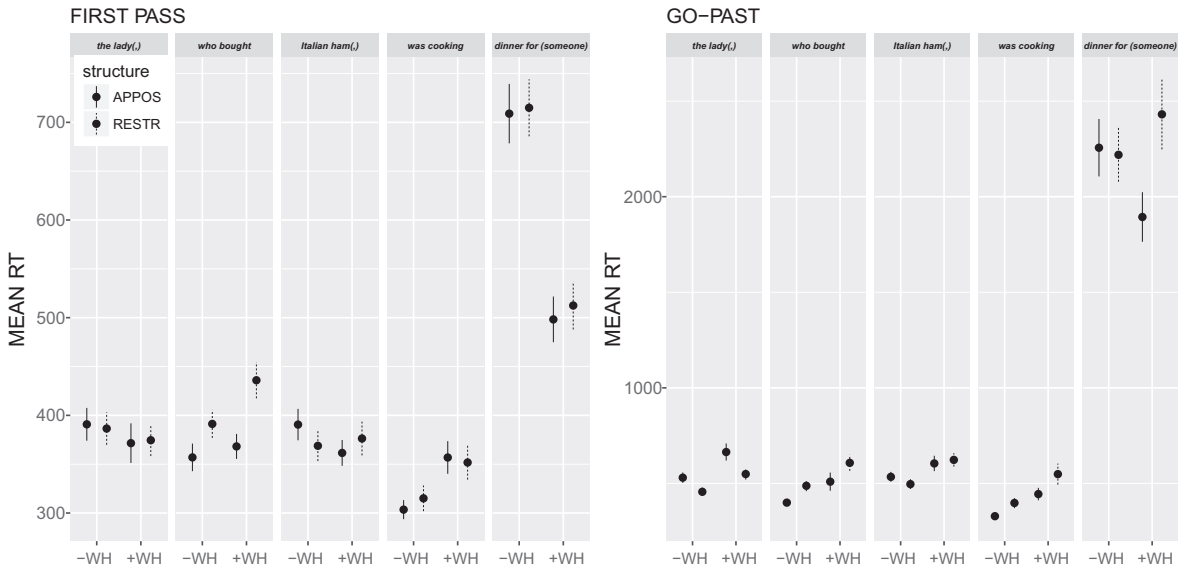

Fig. 2. Mean first-pass and regression path times on the embedded clause region for Experiment 2; error bars represent standard error by participants.

Table 5
Summary of LME coefficient estimates, standard errors (in parentheses) and associated *t*-values for log first-pass reading times, Experiment 2. Bold cells indicate significant values.

	the lady	*t*	*who bought*	*t*	*Italian ham*	*t*	*was cooking*	*t*	*dinner for*	*t*
wh	**−0.07 (0.03)**	**−2.1**	0.06 (0.03)	1.9	−0.02 (0.03)	−0.6	0.07 (0.04)	1.8	**−0.37 (0.04)**	**−8.4**
structure	−0.03 (0.04)	−0.8	**−0.11 (0.03)**	**−3.5**	0 (0.04)	0.0	0 (0.03)	0.1	−0.03 (0.04)	−0.7
wh:structure	−0.03 (0.05)	−0.5	−0.05 (0.05)	−0.9	−0.05 (0.05)	−1.0	0.01 (0.06)	0.2	−0.06 (0.07)	−0.8

Table 6
Summary of LME coefficient estimates, standard errors (in parentheses) and associated *t*-values for log go-past reading times, Experiment 2. Bold cells indicate significant values.

	the lady	*t*	*who bought*	*t*	*Italian ham*	*t*	*was cooking*	*t*	*dinner for*	*t*
wh	**0.17 (0.04)**	**4.7**	**0.16 (0.04)**	**4.4**	**0.11 (0.04)**	**2.5**	0.12 (0.06)	1.8	**−0.12 (0.05)**	**−2.5**
structure	**0.14 (0.04)**	**3.8**	**−0.17 (0.04)**	**−4.3**	0.01 (0.04)	0.3	−0.08 (0.04)	−1.8	**−0.12 (0.05)**	**−2.4**
wh:structure	0.03 (0.07)	0.5	−0.03 (0.06)	−0.5	−0.08 (0.06)	−1.3	0.02 (0.08)	0.2	**−0.26 (0.08)**	**−3.3**

APPOS conditions. This interaction was mirrored by qualitatively similar interactions in total viewing times. The overall pattern of reading times, then, suggests that readers encountered significantly more difficulty at the gap site for *RESTR* conditions, and that this difficulty resulted in significant re-reading of the entire sentence.

With respect to the hypotheses we sketched out at the beginning, we find clear evidence in favor of the differential loss hypothesis, and no reliable evidence in support of the total isolation hypothesis: the difficulty of maintaining an open filler-gap dependency across a relative clause was not statistically modulated by the type of relative clause. This raises the possibility that the filler-gap dependency in the main clause is available to interfere with the processing of the relative clause to a comparable extent for both appositives and restrictives, running counter to the predictions of the total isolation hypothesis. However, it must be acknowledged that there is a numerical pattern of first-pass times at the RC region that aligns with the predictions of the total isolation hypothesis: there were numerically longer first pass times in the *RESTR*, *+WH* condition at this region. However, this pattern was not apparent in the go-past times at this region; in this measure the RT penalty for an open *wh*-dependency was comparable for both appositive and restrictive relative clauses. Overall, we find no evidence that supports the total isolation hypothesis, although we cannot reject this hypothesis completely in light of the pattern of first-pass times at the RC region.

The interaction of *wh* movement and *structure* at the gap site lends support to the differential loss hypothesis. It appears that a recently processed appositive relative clause interferes less with filler integration than does a recently-processed restrictive relative clause, consistent with the view that the lexical and/or syntactic material associated with the appositive relative clause is rendered unavailable once it has been processed.

In addition, we observed reading time slowdowns in the APPOS conditions on the embedded subject region, followed by reading time slowdowns for the RESTR conditions at the beginning of the relative clause region. We suspect that this pattern reflects the effect of the comma on reading times: on the embedded subject, the comma in the APPOS conditions may lead to slower reading times, followed by compensatory speeding after the comma that results in slower reading times for RESTR conditions. Hirotani, Frazier, and Rayner (2006) reported a similar pattern of effects triggered by commas in an eye-tracking-while-reading experiment. Indeed, other studies suggest that reading times are faster immediately following a comma, and that saccades are longer when they cross punctuation than when they do not (Rayner, Kambe, & Duffy, 2000; Ren & Yang, 2010; but c.f. Pynte & Kennedy, 2007; Warren, White, & Reichle, 2009). In light of these findings, one potential worry is that our critical effect might simply reflect overall faster reading times following the commas in the appositive conditions. Inspection of the first-pass times following the appositive and restrictive relative clauses weighs against this alternative interpretation, however. While there is a trend towards faster reading times in the appositive conditions following the relative clause (i.e. on *was cooking*), there is little evidence in the reading times that this post-comma speed-up persists into the critical gap region. In particular, first past reading times at the critical gap region present no clear difference between *APPOS* and *RESTR* conditions, indicating that there is no general appositive advantage in first-pass times that persists into the final region. Moreover, we observe that the first-pass reading times on the final region do not appear to differ substantially between the *+WH RESTR* and *+WH APPOS* conditions. Instead, the difference between these conditions arises in go-past times. This pattern suggests that the difficulty associated with *+WH RESTR* is the result of increased regressions and re-reading in that condition, rather than facilitated processing in the *+WH APPOS* condition. For these reasons we find it unlikely that the effects we observe at the gap site are a direct reflection of the influence of commas on reading times in the *APPOS* conditions.

Lastly, we note that our experiment differed from previous investigations of appositive relative clauses in reading. In our experiment, we used only orthographic cues (commas) to indicate appositive relative clauses, instead of providing both contextual and orthographic cues (as did Gibson, Desmet, Grodner, Watson, & Ko, 2005; Grodner et al., 2005). However, the results of our offline post-test suggest that our manipulation was sufficient to convey the appositive/restrictive contrast. Our participants reliably used commas to distinguish the two classes of relative clause.

One clear weakness with our preferred interpretation of the data is that the critical interaction occurred sentence finally. Thus, we cannot rule out the possibility that this effect is not due to gap-filling per se, as opposed to a wrap-up effect triggered by the end of the sentence. Experiment 3 addresses this possibility.

Experiment 3: eye-tracking while reading

In Experiment 3, we seek to replicate and extend the findings in Experiment 2. We again use eye-tracking while reading to test for an interaction of *wh*-movement and relative clause structure. The critical difference between Experiment 3 and Experiment 2 is the position of the gap in our materials. In Experiment 2, the gap was sentence-final; in Experiment 3, we changed the materials so that all gaps were sentence-medial, followed by a spillover region. The differential loss hypothesis holds that the penalty in the *+WH*, *RESTR* conditions in Experiment 2 results from the

selective interference that the restrictive relative clause creates for the purposes of integrating the filler with its gap. If this is correct, then in Experiment 3 we expect to see our critical interaction earlier in the sentence. If instead the effect reflects differences in wrap-up processing, we expect to see no difference in the timing of the interaction between Experiment 2 and Experiment 3.

Methods

Subjects

37 native speakers of American English were recruited to participate in Experiment 3; of these, 2 participants' data was removed prior to analysis due to unacceptable levels of track loss, leaving a total of 35 participants included in the final analysis. Participants were recruited in undergraduate Psychology classes at the University of Massachusetts, Amherst. Participants gave informed consent and were given course credit in exchange for participation. Completion of the experiment, including calibration of the eye-tracker, took approximately 45 min.

Materials

The 24 critical items from Experiment 2 were modified to create the items for Experiment 3. In all items, the embedded verb was transitive and took an immediately following direct object. In *+WH* conditions, the direct object was the position of the gap site. Following the critical verb and gap site, each item was given a spillover region. The effect of these changes was to place the critical gap region at a sentence-medial position, rather than the sentence-final position in Experiment 2. An example, along with regions of interest for eye-tracking analysis, is given in (6):

6. a. [**RESTR, +WH**]/The butcher/asked who/the lady/who bought/Italian ham/had invited _/to dinner tonight./
   b. [**RESTR, −WH**]/The butcher/asked if/the lady/who bought/Italian ham/had invited/anyone to dinner tonight./
   c. [**APPOS, +WH**]/The butcher/asked who/the lady,/who bought/Italian ham,/had invited/to dinner tonight./
   d. [**APPOS, −WH**]/The butcher/asked if/the lady,/who bought/Italian ham,/had invited _/anyone to dinner tonight./

Underscores mark the position of the gap; the underscore was not presented to participants in the experiment. The full set of items is provided in Appendix B.

The 24 target items were combined with 86 unrelated fillers for a total of 100 sentences. All filler sentences were of comparable length and complexity to the target items, but were a different set of fillers from those used in Experiment 2. All fillers were grammatical, and contained a range of structural and referential ambiguities.

Procedure

The experimental procedure was largely identical to Experiment 2. In Experiment 3, however, 3-point calibration was used. All experimental items fit on a single line of text.

Analysis

Statistical analysis was performed as in Experiment 2.

Results

Offline responses

Participants performed well in the task as indicated by their performance on the comprehension questions. Accuracy on the filler items was 84%, and accuracy rates on experimental items ranged from 87% to 91% across conditions. As in Experiment 2, LME modeling of the responses yielded no significant differences between conditions.

Reading time measures

The mean total viewing times in Experiment 3 are presented in Table 7, and their distribution is presented in Fig. 3. Table 8 presents the results of the LME analysis on log-transformed total viewing times. LME analysis revealed an effect of *wh* and an interaction of *wh* and *structure*, reflecting significantly longer total viewing times in the *+WH*, *RESTR* condition than all other conditions.

Raw first-pass and go-past times for each region are presented in Fig. 4; tables of mean reading times are in Appendix A. Tables 9 and 10 present the results of LME analyses on log-transformed first-pass and go-past times, respectively.

As in Experiment 2, there was an interaction of *wh* and *structure* on go-past times at the gap site. Importantly, we observed this interaction at the embedded gap site (*had invited*), not at the sentence-final region. As in Experiment 2, this interaction was driven by a larger effect of *WH* for *RESTR* conditions than for *APPOS* conditions. Pairwise comparisons at the critical embedded verb region confirmed that the effect of *WH* was more pronounced for *RESTR* conditions ($\bar{x}$ = 0.2, 95% CI: [0.1, 0.3]) than *APPOSITIVE* conditions ($\bar{x}$ = 0, 95% CI: [−0.1, 0.1]).

However, one surprising finding was that the same interaction of *wh* and *structure* was also observed in go-past times on the region immediately preceding the gap site, at the end of the embedded relative clause (*Italian ham*,); we return to this surprising finding in the discussion. With the exception of this interaction, however, reading time effects on the regions preceding the gap site were qualitatively similar to those observed in Experiment 2. At the embedded subject region, we observed significant main effects of *wh* and *structure* on go-past times. At the beginning of the relative clause region (*who bought*), we observed a main effect of *WH* in both first-pass and go-past times, confirming the basic pattern observed in Experiment 2. As in Experiment 2, we saw a trend towards an interaction of *wh* and *structure* in first-pass times at this region (t = 1.8), reflecting greater reading times in the *+WH*, *RESTR* condition. As before, this did not persist into go-past times. We take up this trend, and its cross-experiment consistency, in the general discussion.

A point of difference between the two experiments is that in Experiment 3, we did not observe the same effect of *structure* at the relative clause region that we did in Experiment 2 (viz. slower RTs for *RESTR* conditions). As in Experiment 2, we observed no significant first-pass effects at the end of the relative clause or at the embedded verb region, but we did observe a large effect of *WH* at the sentence final region: −WH conditions were read significantly longer, presumably a reflection of the differing lengths of this region in +WH and −WH conditions. In go-past times, we also saw main effects of *WH* and *structure* at the embedded verb region, and a main effect of *WH* in the sentence final region.

Table 7
Mean total viewing times for Experiment 3. Parentheses represent standard error by participants.

	APPOS, −WH	APPOS, +WH	RESTR, −WH	RESTR, +WH
Total viewing time	4871 (236)	4949 (241)	4576 (199)	5356 (257)

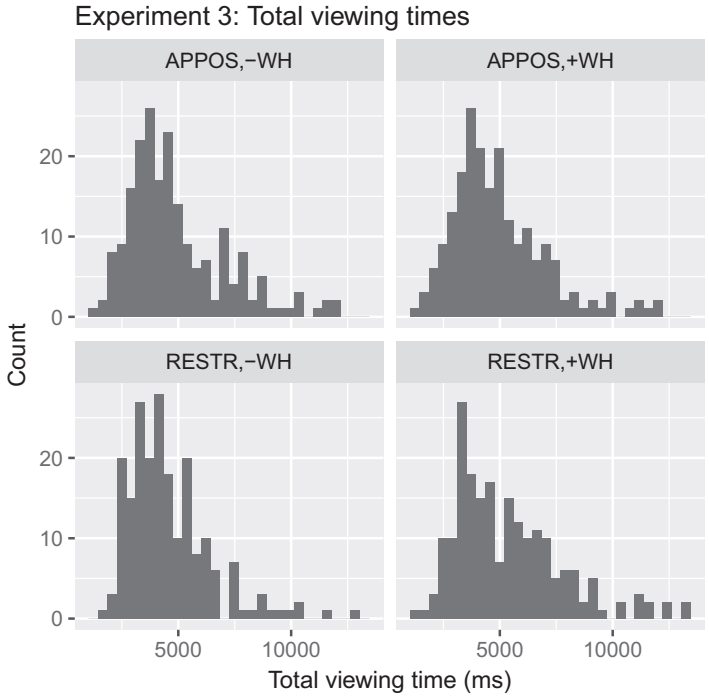

Fig. 3. Histogram showing the distribution of total viewing times by condition in Experiment 3.

Table 8
Summary of LME analyses of log total viewing times, Experiment 3. Bold cells indicate significant coefficients.

	TVT: $\hat{\beta}$	t
wh	**0.08 (0.04)**	**2.29**
structure	−0.01 (0.02)	−0.45
wh:*structure*	**−0.11 (0.05)**	**−2.16**

Discussion

The findings in Experiment 3 accord well with those of Experiment 2. At the point at which the relative clause was introduced (*who bought*), we observed significant slowdowns in first-pass and go-past times in *+WH* conditions. However, we again failed to observe any significant interaction between this slowdown and the structure of the relative clause: as in Experiment 2, there was a limited suggestion of this interaction in first-pass times that was not at all evident in the go-past times.

At the critical gap region (*had invited*), we observed an interaction of *wh* and *structure* in go-past times, driven by a larger slowdown for *+WH* in *RESTR* conditions than in *APPOS* conditions. Critically, this effect occurred earlier in Experiment 3 than in Experiment 2. In Experiment 3, it was observed at the embedded verb region, which hosted the gap site, rather than sentence-finally, as in Experiment 2. The interaction accords with our

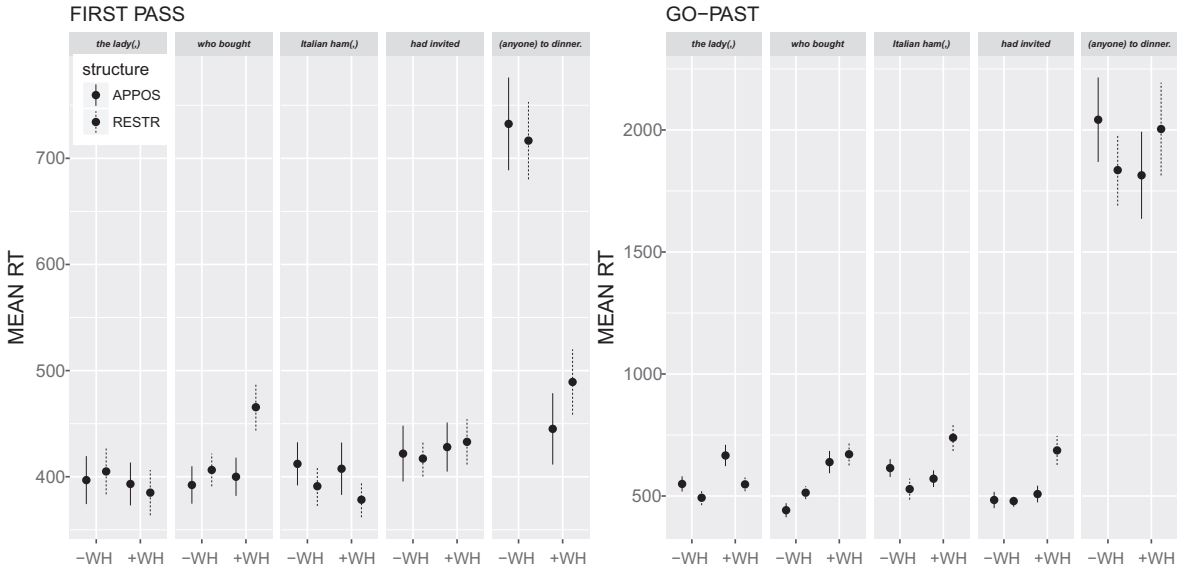

Fig. 4. Mean first-pass and regression path times on the embedded clause region for Experiment 3; error bars represent standard error by participants.

Table 9
Summary of LME analyses of log first-pass reading times, Experiment 3. Bold cells indicate significant coefficients.

	the lady	*t*	*who bought*	*t*	*Italian ham*	*t*	*had invited*	*t*	*to dinner*	*t*
wh	−0.05 (0.03)	−1.5	**0.08 (0.03)**	**2.3**	−0.04 (0.04)	−1.0	0.02 (0.04)	0.5	**−0.48 (0.05)**	**−9.1**
structure	−0.03 (0.03)	−1.0	−0.06 (0.04)	−1.6	0.03 (0.03)	0.7	0 (0.04)	0.0	−0.03 (0.05)	−0.6
wh:structure	0.02 (0.07)	0.3	−0.12 (0.07)	−1.8	0.01 (0.08)	0.2	0 (0.07)	0.0	−0.1 (0.08)	−1.3

Table 10
Summary of LME analyses of log go-past times, Experiment 3. Bold cells indicate significant coefficients.

	the lady	*t*	*who bought*	*t*	*Italian ham*	*t*	*had invited*	*t*	*to dinner*	*t*
wh	**0.16 (0.05)**	**3.6**	**0.25 (0.05)**	**5.2**	0.05 (0.05)	1.0	**0.12 (0.04)**	**2.7**	**−0.17 (0.07)**	**−2.3**
structure	**0.14 (0.05)**	**2.9**	−0.06 (0.05)	−1.4	−0.01 (0.04)	−0.3	**−0.13 (0.05)**	**−2.4**	0 (0.05)	0.1
wh:structure	0.06 (0.07)	0.9	0.06 (0.08)	0.8	**−0.28 (0.09)**	**−3.1**	**−0.18 (0.07)**	**−2.5**	−0.11 (0.12)	−0.9

interpretation that this interaction reflects processes associated with retrieving and integrating the filler at its gap site. Because this effect reliably localizes to the position of the gap in the sentence, we interpret this as evidence that the content of the restrictive relative clause is more available to create retroactive interference with the retrieval of the filler than the content of the appositive relative clause, supporting the differential loss hypothesis.

However, one surprising finding was that we also observed an interaction of *wh* and *structure* in go-past times at the end of the relative clause (*Italian ham*), which preceded the critical gap site. This effect was unexpected, and the source of this interaction remains unclear. To better understand this effect, we conducted a post hoc split-half analysis of the experiment. The results of this analysis are presented in Appendix C. This split-half analysis revealed two findings. One, the critical interaction at the gap site was numerically present in both halves of the experiment. Two, the unexpected interaction at the end of the relative clause was present only in the first half of the experiment. Unfortunately, we can only speculate as to the source of this interaction. It may reflect difficulty readers experience when deciding where to terminate the restrictive relative clause in the *+WH* conditions. Critically, however, the interaction at the gap site survived even after this early effect disappeared, which supports our interpretation of the interaction in go-past times as a reflection of routine gap-driven processing.

The most reliable pattern that is consistent across Experiments 2 and 3 is the interaction of *wh* and *structure* at the gap site. However, we note that Experiment 3 like Experiment 2 presented a numerical trend towards an interaction of *wh* and *structure* in first-pass times on the beginning of the relative clause region *who bought*. This pattern was not statistically reliable in either experiment. However, it is notable that interaction was in the direction predicted by the total isolation hypothesis; for this reason, we cannot dismiss this trend out of hand. We take up the ramifications of this numerical trend in the general discussion.

General discussion

Across three experiments, we investigated the degree to which an open filler-gap dependency interacted with restrictive and appositive relative clauses that intervened between the filler and its gap site. In Experiment 1 we observed that the presence of a filler-gap dependency significantly decreased offline acceptability ratings compared to controls without such a dependency. However, this penalty was smaller when the filler-gap dependency spanned an appositive relative clause than when it spanned a restrictive relative clause. This finding extends the conclusions of Dillon et al. (2014) concerning the limited interactivity of at-issue and not-at-issue content, and shows that it holds for appositive relative clauses in addition to adnominal appositives.

In Experiments 2 and 3, we used eye-tracking-while-reading to determine how this effect plays out in the course of incremental processing. The results of both experiments support the conclusion that appositive relative clauses interfere less with the processing of a filler-gap dependency than do restrictive relative clauses during eye-tracking-while-reading. Specifically, in both Experiments 2 and 3 we observed an interactive pattern in go-past times at the gap position: there was a slowdown at gap site for conditions with an open filler-gap dependency, but this slowdown was smaller when there was a preceding appositive than when there was a preceding restrictive relative clause. In addition, we observed a significant interaction in total viewing times for both experiments: difficulty at the gap site in the *+WH*, *RESTR* conditions appeared to trigger significant re-reading of the sentence, a pattern that was not observed for *+WH*, *APPOS*. This finding aligns well with the view that once comprehenders have finished processing the appositive relative clause, its form is less available in short-term memory than an otherwise identical restrictive relative clause, in line with the differential loss hypothesis.

At the beginning of the relative clause, we observed only main effects of relative clause type and the presence of a filler-gap dependency on reading measures. Importantly, readers spent more time reading a relative clause when they were also maintaining an open filler-gap dependency. The magnitude of this slowdown was comparable for appositive and restrictive relative clauses in go-past times. Overall, our data support the following three claims: (i) integration of the filler at the gap site is more difficult when readers have recently processed a restrictive relative clause than when they have recently processed an appositive relative clause, (ii) this difficulty results in significant re-reading of the entire sentence, and (iii) we failed to find any reliable evidence that an open filler-gap dependency interferes with the initial processing of a relative clause differentially for appositive and restrictive relative clauses.

On balance, this pattern of results is more consistent with the differential loss hypothesis than it is with total isolation hypothesis. The central prediction of the total isolation hypothesis is that the independence between not-at-issue appositives and at-issue restrictives should be in evidence from the earliest point the appositive is encountered. We found no reliable evidence to support this prediction. Instead, at the initial relative clause region *who bought* go-past times were slowed by an open filler-gap dependency to a comparable extent for appositive and restrictive relative clauses alike. Although the pattern in go-past times seems relatively clear evidence that an open filler-gap dependency taxes both appositives and restrictives at this critical early region, the pattern of RTs in first-pass reading at this region offers a more

mixed picture. In both experiments we did observe a non-reliable trend towards an interaction in first-pass reading times, such that only restrictive relative clauses show any discernable difficulty on the RC region itself. Although it did not reach significance in either experiment, this pattern is consistent with the total isolation hypothesis. For this reason we cannot reject this hypothesis as entirely inconsistent with the data; it is possible that the effect may have just been too small to detect in our experiments, and that our failure to find the effect is in essence a Type II error. As mentioned above, however, this pattern was not observed in other measures at the beginning of the relative clauses: in go-past times, the penalty for an open filler-gap dependency was comparable across both types of restrictive clauses. This pattern suggests that even if there were a small interaction at the beginning of the relative clause, it is relatively small and short-lived. For this reason we believe it unlikely that this effect is the primary source of the limited interactivity between not-at-issue and at-issue content that we observe in our offline ratings and total viewing time data. Overall, then, we do not believe the total isolation hypothesis to be a plausible explanation of our findings.

In contrast, the pattern of reading times at the gap site in Experiments 2 and 3 lends fairly direct support to the differential loss hypothesis. Recall that on this hypothesis, the limited interactivity between not-at-issue appositives and their host clause obtains because the parser does not maintain an active syntactic representation of the appositive relative clause once it has finished processing it; in contrast, the at-issue restrictive relative clauses are kept active throughout the processing of the entire sentence. Because of this, appositive relative clauses are predicted to interfere relatively little with the processing of the host clause after they have been processed. Specifically, we assume that the syntactic features of the relative pronoun *who*, which overlap with the syntactic features of the filler, interfere with the retrieval of the appropriate filler at the gap site (see Obata et al., 2010; Wagers, 2008). If the syntactic form of the appositive relative clause is no longer available, then the material inside the appositive relative should be less available to create this interference, effectively isolating the material inside the appositive from the matrix clause and thereby making integration of the true filler at this position easier and less prone to error in the appositive conditions.

We have interpreted our results to suggest that the syntactic form of the appositive is lost relatively rapidly, minimizing its ability to interfere with subsequent processing. It is interesting to compare this with the observation that comprehenders lose access to the syntactic form of a sentence fairly rapidly. This often results in impaired recognition (Sachs, 1967) and recall (Jarvella, 1971; Lombardi & Potter, 1992; Potter, 1990; Potter & Lombardi, 1990, 1998) for the surface syntactic form of a sentence after even a very short delay, although some studies have demonstrated non-trivial recognition and recall of syntactic form at even long delays (Bates, Masling, & Kintsch, 1978; Cleary & Langley, 2007; Gurevich, Johnson, & Goldberg, 2010; Sachs, 1974). The cause of this phenomenon remains up for debate. It could be that comprehenders rapidly recode the sentence into a more durable long-term format, thereby losing access to the surface form (e.g. Bever & Garrett, 1974). An alternative possibility is that comprehenders routinely encode the surface syntactic form of the sentences they process, but lose access to those encodings due to the interference created by subsequent linguistic material (Wagers, 2014). Whatever the source, the differential loss hypothesis endorses the claim that this short-term 'forgetting' of the syntactic form of a sentence occurs during incremental sentence comprehension. The present data suggest that the syntactic form of an appositive clause is less available after it has been processed, perhaps because it forms a distinct discourse unit in a way that the restrictive relative clause does not. At present we have marshaled support for this view by investigating the degree to which the syntactic form of the relative clause is available to create retroactive interference with syntactic processing; however this view makes clear predictions about the availability of the surface form of the appositive relative clause at various points during comprehension. We leave testing of these claims to future work.

Turning to our finding that there is +*WH* penalty at the beginning of the relative clause, we note that this is compatible with most existing theories of syntactic complexity. The finding that this penalty is not modulated by the status of the relative clause is also compatible with leading theories, as they do not generally model the difference between appositive and restrictive relative clauses (Grodner et al., 2005). However, the particular explanation of this effect varies between theories. For example, the DLT holds that this effect reflects the additional storage cost incurred by opening a second filler-gap dependency inside the relative clause structure (Grodner et al., 2002; Warren & Gibson, 2002). In contrast, the cue-based parsing model offered by Lewis and Vasishth (2005) posits that proactive retrieval interference from the matrix clause filler-gap dependency interferes with the completion of the local filler-gap dependency inside the relative clause (Obata et al., 2010). A third possibility is that the difficulty encountered at the embedded relative clause reflects the comprehender's relative inexperience with these structures, creating high surprisal values at this point in the sentence (e.g. Levy, 2008; MacDonald, 2013). Lastly, it is possible that the early main effects of +WH could simply reflect difference in the frequency with which the embedding verbs we used take +WH complements. The present results do not distinguish these explanations of the difficulty associated with the initial processing the relative clause structure. However, the present results do suggest that however this effect is to be explained, the appositive/restrictive relative clause distinction does not strongly modulate the difficulty associated with these structures.

The role of punctuation in the appositive/restrictive contrast

Our claim that the syntactic representations of appositive and restrictive relative clauses were differentially available in the course of parsing was grounded in their status as independent speech acts. Put bluntly, this position holds that the appositive is to some degree its own sentence that is linearly positioned within another sentence, and this is why the appositive's surface form is lost once it has been processed. More generally, this explanation of our results claims that the differential availability of appositives and restrictives is a reflection of the mapping between units of perceptual organization and units of discourse organization.

However, we can envision at least one alternative explanation for the difference we observe between appositives and restrictives. In our experiments, appositive relative clauses were always indicated orthographically by commas. It is thought that readers treat commas as a reliable indicator of substantial intonational breaks (Chafe, 1988; Steinhauer, 2003; Steinhauer & Friederici, 2001; Kerkhofs, Vonk, Schriefers, & Chwilla, 2007). For example, Steinhauer and Friederici (2001) found that commas yielded similar behavioral and electrophysiological responses in reading paradigms as did intonational phrase boundaries in auditorily-presented stimuli. If readers in our experiment assigned implicit intonational phrase boundaries at the beginning and end of the appositive relative clauses, then it is possible that the difference in availability (or accessibility) of the appositive follows from the prosodic differences between the two, rather than the appositive's status as an independent speech act.

This alternative explanation of the results aligns well with claims made by Frazier, Carlson, and Clifton (2006), who suggested that prosodic structure rather directly determined the grouping of

linguistic units in memory. More specifically, they suggested that prosodic units provide a 'skeleton' within which linguistic processing is performed. There is good evidence for this view outside of the domain of linguistic processing. For example, performance in serial recall tasks is improved when metrical structure is imposed on the study lists (Reeves, Schmauder, & Morris, 2000). Findings like this raise the possibility that prosodic structure and accessibility/availability in memory may be tightly linked. We find the alternative explanation for our results in terms of prosody no less interesting than the explanation of the results in terms of the not-at-issue/at-issue distinction, though it would not explain differences in the interpretation of other not-at-issue structures such as expressives (Frazier et al., 2015). Further research is necessary to determine to what degree the effects reported here reflect the distinct prosodic or interpretive status of the appositive relative clause.

Conclusion

Across three experiments, we found that material inside of appositive relative clauses interferes less with filler-gap processing than does restrictive relative clauses. In particular, it appears that restrictive relative clauses create more interference for retrieval of the filler at a gap position than do appositive relative clauses. This pattern suggests that the syntactic content of restrictive relative clauses is more available in memory than comparable appositive relative clauses after they have been processed. We suggest that this reflects the special role that appositive relative clauses play in the discourse: as independent speech acts, the processor need not actively maintain the surface syntactic or phonological form of an appositive. It appears that the processor may lose access to the surface form of independent discourse units even if those units appear internal to another sentence, as in the case of appositive relative clauses.

Acknowledgments

We would like to thank Jesse Harris and Dave Kush for helpful discussion of the current project, as well as Kristen Syrett and two anonymous reviewers for helpful comments on an earlier draft of this work. All errors that remain are our own.

A. Raw reading times and regressions out for Experiments 2 and 3

See Tables A.1–A.4.

B. Experimental stimuli

Experimental stimuli for Experiments 1 (#1–20) and for Experiment 2 (#1–24):

1. The butcher asked who/if the lady(,) who bought Italian ham (,) was cooking dinner for (her guests).
2. The interviewer wondered who/if the expert(,) who advises the hospital(,) took advice from (anyone).
3. The administrator wanted to know who/if the student(,) who was expelled(,) complained angrily about (Harry).
4. The conductor asked who/if the violinist(,) who was very upset(,) came to the rehearsal with/early.
5. The homeowner asked who/if the carpenter(,) who fixed the roof(,) bought the materials from (Home Depot) in the morning.
6. The choreographer didn't know who/if the dancer(,) who danced clumsily(,) liked to practice with (the teacher).
7. The investigator wondered who/if the victim(,) who was recovering slowly(,) said she would sue (the paramedics).
8. The reporter couldn't remember who/if the auditor(,) who works for the town(,) gave a memo to (the mayor).
9. The comedian didn't know who/if the woman(,) who heckled him mercilessly(,) was sitting next to (his friend).
10. The oil expert wondered who/if the businessman(,) who hated fossil fuels(,) sells solar panels to (many people).
11. The judge asked who/if the defendant(,) who was accused of murder(,) sought legal advice from (a well-known lawyer).
12. The contractor wondered who/if the gardener(,) who loves cherry trees(,) did the landscaping for (the sheriff).
13. The applicant asked who/if the man(,) who was conducting the interviews(,) would give the job to (the younger candidate).
14. The little girl wondered who/if the boy(,) who was ignoring her(,) was writing letters to (his girlfriend).
15. The lobbyist knew who/if the senator(,) who voted for higher taxes(,) promised campaign jobs to (his family).
16. The author asked who/if the bookstore owner(,) who organized the book-signing(,) gave an invitation to (his friend).
17. The customer didn't remember who/if the butcher(,) who sold organic beef(,) buys his meat from (local farmers).
18. The neighbor wondered who/if the local prankster(,) who was very clever(,) would play a prank on (the school teacher) next.
19. The nurse asked who/if the doctor(,) who treated the crash victim(,) prescribes pain medication to (everyone).
20. The reporter wanted to know who/if the hacker(,) who coded the software(,) is designing a new virus for (the government).
21. The teacher asked who/if the student(,) who was failing math(,) got outside tutoring from (anyone).
22. Billy forgot who/if the neighbor(,) who moved from Indiana (,) became good friends with (Mary).
23. Pauline definitely knows who/if the bully(,) who beat up George(,) is going to pick on (Harry) next.
24. The barista wondered who/if the customer(,) who ordered a decaf latte(,) was texting with (her friend).

Experimental stimuli for Experiment 3:

1. The butcher asked who/if the customer(,) who bought Italian sausage(,) had invited (anyone) to dinner tonight.
2. The administrator wondered who/if the doctor(,) who manages the hospital(,) had treated (anyone) in the ER.
3. The principal wanted to know who/if the student(,) who was expelled(,) had threatened (anyone) with a knife.

Table A.1
Mean first-pass reading times for Experiment 2. Parentheses indicate standard error by participants.

	the lady	*who bought*	*Italian ham*	*was cooking*	*dinner for*
APPOS, −WH	391 (17)	357 (14)	391 (16)	303 (10)	709 (30)
APPOS, +WH	372 (20)	368 (13)	362 (13)	357 (17)	498 (23)
RESTR, −WH	387 (17)	391 (14)	369 (16)	315 (13)	715 (29)
RESTR, +WH	375 (16)	436 (18)	376 (17)	352 (18)	512 (25)

Table A.2
Mean go-past times for Experiment 2. Parentheses indicate standard error by participants.

	the lady	*who bought*	*Italian ham*	*was cooking*	*dinner for*
APPOS, −WH	530 (27)	400 (16)	534 (26)	329 (16)	2256 (150)
APPOS, +WH	664 (45)	509 (47)	605 (40)	444 (33)	1894 (130)
RESTR, −WH	456 (22)	488 (28)	496 (26)	398 (27)	2219 (141)
RESTR, +WH	549 (29)	608 (42)	623 (35)	549 (55)	2431 (183)

Table A.3
Mean first-pass reading times for Experiment 3. Parentheses indicate standard error by participants.

	the lady	*who bought*	*Italian ham*	*had invited*	*to dinner*
APPOS, −WH	397 (23)	392 (18)	412 (20)	422 (26)	732 (44)
APPOS, +WH	393 (20)	400 (18)	408 (25)	428 (23)	445 (34)
RESTR, −WH	405 (22)	406 (15)	391 (19)	417 (17)	717 (36)
RESTR, +WH	385 (21)	466 (22)	378 (17)	433 (21)	489 (31)

Table A.4
Mean go-past times for Experiment 3. Parentheses indicate standard error by participants.

	the lady	*who bought*	*Italian ham*	*had invited*	*to dinner*
APPOS, −WH	549 (31)	442 (29)	615 (37)	484 (33)	2042 (173)
APPOS, +WH	666 (44)	639 (46)	571 (34)	508 (34)	1814 (178)
RESTR, −WH	493 (30)	514 (26)	528 (43)	479 (23)	1836 (146)
RESTR, +WH	548 (28)	671 (46)	739 (55)	687 (60)	2004 (190)

4. The conductor asked who/if the violinist(,) who was very upset(,) had talked to (anyone) before coming in.
5. The homeowner asked who/if the contractor(,) who fixed the roof(,) would hire (someone) for the next job.
6. The choreographer didn't know who/if the dancer(,) who danced clumsily(,) had run into (anyone) during practice.
7. The cop wondered who/if the victim(,) who was recovering slowly(,) would blame (the police) for the accident.
8. I couldn't remember who/if the reporter(,) who does the morning news(,) had mentioned (me) in the big story.
9. The singer didn't know who/if the audience member(,) who laughed out loud(,) was heckling (him) during the show.
10. The oilman wondered who/if the businessman(,) who hated fossil fuels(,) had fired (someone) from his company.
11. The judge asked who/if the lawyer(,) who represented the defendant(,) was working with (someone) on the case.
12. The contractor wondered who/if the gardener(,) who loves tulips(,) had met (anyone) at the nursery.
13. The applicant asked who/if the employee(,) who conducted the interviews(,) would hire (someone) if he could.
14. The little girl wondered who/if the boy(,) who was ignoring her(,) had played with (anyone) during recess.
15. The lobbyist knew who/if the senator(,) who voted for higher taxes(,) had bribed (someone) during the election.
16. The author asked who/if the young fan(,) who organized the book-signing(,) had invited (Mary) to the event.
17. Joan didn't remember who/if the butcher(,) who sold organic beef(,) had hired (someone) to deliver his meat.
18. Alex wondered who/if the teenager(,) who was a clever prankster(,) would target (John) for his next prank.
19. The nurse asked who/if the doctor(,) who treated the crash victim(,) had examined (everyone) very thoroughly.
20. The reporter wanted to know who/if the hacker(,) who coded the software(,) had attacked (the NSA) with a virus.
21. The teacher asked who/if the student(,) who was failing math(,) could talk to (anyone) for some help.
22. Billy forgot who/if the neighbor(,) who moved from Indiana (,) was dating (Susan) this time last year.
23. Pauline definitely knows who/if the bully(,) who beat up George(,) was harassing (Paul) this week.
24. The barista wondered who/if the customer(,) who ordered a decaf latte(,) was meeting (someone) in the cafe .

C. Split-half analysis of Experiment 3

The pattern of reading times at the end of the relative clause (e.g. *Italian Ham*) suggests that the interaction on this region reported in the main text is driven by the first half of the experiment. LME analysis of this region reveals a significant interaction only in the first half of the experiment; it is no longer significant in the second half, and inspection of the means suggests that the effect of *WH* in the *RESTR* conditions has been eliminated. In contrast, the pattern of means at the critical region (e.g. *had invited*) remains stable across both halves of the experiment, although we note that it only reaches significance in the second half (see Tables C.1–C.4).

Table C.1
Mean go-past reading times for the first 12 experimental items presented in Experiment 3. Parentheses indicate standard error by participants.

	the lady	*who bought*	*Italian ham*	*had invited*	*to dinner*
APPOS, −WH	579 (39)	421 (34)	639 (45)	493 (43)	2199 (195)
APPOS, +WH	728 (61)	751 (76)	644 (74)	524 (40)	1792 (164)
RESTR, −WH	519 (51)	492 (43)	505 (43)	461 (26)	2061 (206)
RESTR, +WH	563 (40)	758 (85)	892 (113)	624 (70)	2392 (218)

Table C.2
Mean go-past reading times for the last 12 experimental items presented in Experiment 3. Parentheses indicate standard error by participants.

	the lady	*who bought*	*Italian ham*	*had invited*	*to dinner*
APPOS, −WH	530 (42)	454 (31)	529 (45)	459 (36)	1891 (228)
APPOS, +WH	610 (47)	560 (50)	471 (28)	481 (46)	1824 (275)
RESTR, −WH	483 (38)	553 (44)	514 (61)	496 (33)	1598 (161)
RESTR, +WH	539 (35)	604 (43)	590 (52)	716 (74)	1592 (205)

Table C.3
Summary of LME analyses of log go-past times on the end of the relative clause and critical region for the first 12 experimental items in Experiment 3.

	Italian Ham: $\hat{\beta}$	*t*	*Had invited*: $\hat{\beta}$	*t*
wh	0.12 (0.07)	1.63	**0.14 (0.05)**	**2.63**
structure	−0.02 (0.07)	−0.28	−0.04 (0.06)	−0.74
wh:*structure*	**−0.37 (0.14)**	**−2.70**	−0.13 (0.1)	−1.33

Table C.4
Summary of LME analyses of log go-past times on the end of the relative clause and critical region for the second 12 experimental items in Experiment 3.

	Italian Ham: $\hat{\beta}$	*t*	*Had invited*: $\hat{\beta}$	*t*
wh	−0.02 (0.07)	−0.30	0.09 (0.06)	1.40
structure	−0.02 (0.06)	−0.38	**−0.22 (0.07)**	**−3.02**
wh:*structure*	−0.19 (0.11)	−1.64	**−0.21 (0.1)**	**−2.16**

References

Amaral, P., Roberts, C., & Smith, E. A. (2007). Review of the logic of conventional implicatures by Chris Potts. *Linguistics and Philosophy, 30*, 707–749. http://dx.doi.org/10.1007/s10988-008-9025-2.

AnderBois, S., Brasoveanu, A., & Henderson, R. (2011). Crossing the appositive/at-issue meaning boundary. *Proceedings of SALT, 20*, 328–346. http://dx.doi.org/10.3765/salt.v20i0.2551.

AnderBois, S., Brasoveanu, A., & Henderson, R. (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics, 32*, 93–138. http://dx.doi.org/10.1093/jos/fft014.

Arnold, D. (2007). Non-restrictive relatives are not orphans. *Journal of Linguistics, 43*, 271–309. http://dx.doi.org/10.1017/S0022226707004586.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software, 67*, 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Bates, E., Masling, M., & Kintsch, W. (1978). Recognition memory for aspects of dialogue. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 187–197. http://dx.doi.org/10.1037/0278-7393.4.3.187.

Bever, T. G., & Garrett, M. F. (1974). *The Psychology of Language: an Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw-Hill.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B, 26*, 211–252.

Chafe, W. (1988). Punctuation and the prosody of written language. *Written Communication, 5*, 395–426. http://dx.doi.org/10.1177/0741088388005004001.

Cleary, A. M., & Langley, M. M. (2007). Retention of the structure underlying sentences. *Language and Cognitive Processes, 22*, 614–628. http://dx.doi.org/10.1080/01690960601049438.

Dillon, B., Clifton, C., Jr., & Frazier, L. (2014). Pushed aside: Parentheticals, memory and processing. *Language, Cognition and Neuroscience, 29*, 483–498. http://dx.doi.org/10.1080/01690965.2013.866684.

Fanselow, G., & Frisch, S. (2006). Effects of processing difficulty on judgments of acceptability. In G. Fanselow, C. Féry, M. Schlesewsky, & R. Vogel (Eds.), *Gradience in grammar: Generative perspectives* (pp. 291–316). New York, NY: Oxford University Press.

Fedorenko, E., Woodbury, R., & Gibson, E. (2013). Direct evidence of memory retrieval as a source of difficulty in non-local dependencies in language. *Cognitive Science, 37*, 378–394. http://dx.doi.org/10.1111/cogs.12021.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research, 30*, 321–338. http://dx.doi.org/10.1023/A:1010447102554.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language, 47*, 250–272.

Frazier, L., Carlson, K., & Clifton, C. Jr., (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*, 244–249. http://dx.doi.org/10.1016/j.tics.2006.04.002.

Frazier, L., & Clifton, C. Jr., (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes, 4*, 93–126. http://dx.doi.org/10.1080/01690968908406359.

Frazier, L., Dillon, B., & Clifton, C. Jr., (2015). A note on interpreting damn expressives: Transferring the blame. *Language and Cognition, 7*, 1–14. http://dx.doi.org/10.1017/langcog.2014.31.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291–325. http://dx.doi.org/10.1016/0010-0277(78)90002-1.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, United Kingdom: Cambridge University Press.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown* (Doctoral dissertation). Carnegie Mellon University. Retrieved from <http://dl.acm.org/citation.cfm?id=124025>.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76. http://dx.doi.org/10.1016/S0010-0277(98)00034-1.

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

Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics, 16*, 313–353. http://dx.doi.org/10.1515/cogl.2005.16.2.313.

Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax, 7*, 55–78. http://dx.doi.org/10.1111/j.1368-0005.2004.00065.x.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1411–1423. http://dx.doi.org/10.1037/0278-7393.27.6.1411.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*, 97–114. http://dx.doi.org/10.1016/j.jml.2004.02.003.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1304–1321. http://dx.doi.org/10.1037/0278-7393.32.6.1304.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13*, 425–430. http://dx.doi.org/10.1111/1467-9280.00475.

Grodner, D., Gibson, E., & Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language, 46*, 267–295. http://dx.doi.org/10.1006/jmla.2001.2808.

Grodner, D., Gibson, E., & Watson, D. (2005). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition, 95*, 275–296. http://dx.doi.org/10.1016/j.cognition.2004.01.007.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12*, 175–204.

Gurevich, O., Johnson, M. A., & Goldberg, A. E. (2010). Incidental verbatim memory for language. *Language and Cognition, 2*, 45–78. http://dx.doi.org/10.1515/langcog.2010.003.

Harris, J. A., & Potts, C. (2009). Perspective-shifting with appositives and expressives. *Linguistics and Philosophy, 32*, 523–552. http://dx.doi.org/10.1007/s10988-010-9070-5.

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language, 54*, 425–443. http://dx.doi.org/10.1016/j.jml.2005.12.001.

Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes, 26*, 376–405. http://dx.doi.org/10.1080/01690965.2010.492642.

Hofmeister, P., Jaeger, T. F., Arnon, I., Sag, I. A., & Snider, N. (2013). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes, 28*, 48–87. http://dx.doi.org/10.1080/01690965.2011.572401.

Hofmeister, P., & Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology, 5*, 1237. http://dx.doi.org/10.3389/fpsyg.2014.01237.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior, 10*, 409–416. http://dx.doi.org/10.1016/S0022-5371(71)80040-3.

Jasinskaja, K. (2016). Not at issue any more Retrieved from: <http://dslc.phil-fak.uni-koeln.de/sites/dslc/katja_files/jasinskaja_any_more.pdf>.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159–201. http://dx.doi.org/10.1080/016909600386084.

Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. (2007). Discourse, syntax, and prosody: The brain reveals an immediate interaction. *Journal of Cognitive Neuroscience, 19*, 1421–1434. http://dx.doi.org/10.1162/jocn.2007.19.9.1421.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition, 2*, 15–47. http://dx.doi.org/10.1016/0010-0277(72)90028-5.

Kush, D., Johns, C. L., & Van Dyke, J. A. (2015). Identifying the role of phonology in sentence-level reading. *Journal of Memory and Language, 79*, 18–29. http://dx.doi.org/10.1016/j.jml.2014.11.001.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests in linear mixed effects models [Computer Software]*. Retrieved from <https://CRAN.R-project.org/package=lmerTest>.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177. http://dx.doi.org/10.1016/j.cognition.2007.05.006.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25*, 93–115. http://dx.doi.org/10.1007/BF01708421.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375–419. http://dx.doi.org/10.1207/s15516709cog0000_25.

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language, 31*, 713–733.

Loock, R. (2007). Appositive relative clauses and their functions in discourse. *Journal of Pragmatics, 39*, 336–362. http://dx.doi.org/10.1016/j.pragma.2006.02.007.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*, 226. http://dx.doi.org/10.3389/fpsyg.2013.00226.

McElree, B. (2006). Accessing recent events. In B. Ross (Ed.). *Psychology of learning and motivation* (Vol. 46, pp. 155–200). San Diego, CA: Elsevier Academic Press.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*, 67–91. http://dx.doi.org/10.1016/S0749-596X(02)00515-6.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.). *Handbook of mathematical psychology* (Vol. 2, pp. 419–491). New York, NY: John Wiley & Sons.

Obata, M., Lewis, R. L., Epstein, S., Bartek, B., & Boland, J. (2010). Featural analysis and short-term memory retrieval in on-line parsing: Evidence for syntactic, but not phonological, similarity-based interference. In Y. Fainleib, N. LaCara, & Y. Park (Eds.). *Proceedings of NELS 41: Conference of the North East Linguistics Society* (Vol. 2, pp. 69–82). Amherst, MA: University of Massachusetts Graduate Student Linguistics Association.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research, 22*, 407–428. http://dx.doi.org/10.1016/j.cogbrainres.2004.09.012.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics, 12*, 601–638. http://dx.doi.org/10.1016/0378-2166(88)90050-1.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29*, 633–654.

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language, 38*, 265–282.

Potts, C. (2005). *The logic of conventional implicatures*. Oxford, United Kingdom: Oxford University Press.

Potts, C. (2015). Presupposition and implicature. In S. Lappin & C. Fox (Eds.). *The handbook of contemporary semantic theory* (Vol. 2, pp. 168–202). Oxford, United Kingdom: Oxford University Press.

Pynte, J., & Kennedy, A. (2007). The influence of punctuation and word class on distributed processing in normal reading. *Vision Research, 47*, 1215–1227. http://dx.doi.org/10.1016/j.visres.2006.12.006.

R Core Team (2015). *R: A language and environment for statistical computing [computer software]*. Vienna, Austria: R Foundation for Statistical Computing.

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology, 53A*, 1061–1080. http://dx.doi.org/10.1080/713755934.

Redeker, G. (2006). Discourse markers as attentional cues at discourse transitions. In K. Fischer (Ed.). *Approaches to discourse particles: Studies in pragmatics* (Vol. 1, pp. 339–358). Amsterdam, The Netherlands: Elsevier.

Reeves, C., Schmauder, A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1638–1654. http://dx.doi.org/10.1037/0278-7393.26.6.1638.

Ren, G. Q., & Yang, Y. (2010). Syntactic boundaries and comma placement during silent reading of Chinese text: Evidence from eye movements. *Journal of Research in Reading, 33*, 168–177. http://dx.doi.org/10.1111/j.1467-9817.2009.01406.x.

Resnik, P. (1992). Left-corner parsing and psychological plausibility. *Proceedings of the 14th conference on computational linguistics* (Vol. 1, pp. 191–197). Stroudsburg, PA: Association for Computational Linguistics.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics, 5*(1), 1–69. http://dx.doi.org/10.3765/sp.5.6.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Attention, Perception, & Psychophysics, 2*, 437–442. http://dx.doi.org/10.3758/BF03208784.

Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory & Cognition, 2*, 95–100. http://dx.doi.org/10.3758/BF03197498.

Schlenker, P. (2010). Supplements within a unidimensional semantics I: Scope. In M. Aloni, V. Kimmelman, F. Roelofsen, G. Sassoon, K. Schulz, & M. Westera (Eds.), *Logic, language and meaning* (pp. 74–83). Berlin, Germany: Springer.

Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, IL: University of Chicago Press.

Sprouse, J. (2008). The differential sensitivity of acceptability judgments to processing effects. *Linguistic Inquiry, 39*, 686–694. http://dx.doi.org/10.1162/ling.2008.39.4.686.

Sprouse, J., Fukuda, S., Ono, H., & Kluender, R. (2011). Reverse island effects and the backward search for a licensor in multiple wh-questions. *Syntax, 14*, 179–203. http://dx.doi.org/10.1111/j.1467-9612.2011.00153.x.

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics. Pragmatics* (Vol. 9, pp. 315–332). New York, NY: New York Academic Press.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition, 116*, 71–86. http://dx.doi.org/10.1016/j.cognition.2010.04.002.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language, 86*, 142–164. http://dx.doi.org/10.1016/S0093-934X(02)00542-4.

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research, 30*, 267–295. http://dx.doi.org/10.1023/A:1010443001646.

Syrett, K., & Koev, T. (2015). Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics, 32*, 525–577. http://dx.doi.org/10.1093/jos/ffu007.

Tonhauser, J. (2011). Diagnosing (not-) at-issue content. In E. Bogal-Allbritten (Ed.). *Proceedings of semantics of under-represented languages (SULA)* (Vol. 6, pp. 239–254). Amherst, MA: University of Massachusetts Graduate Student Linguistics Association.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*, 285–316. http://dx.doi.org/10.1016/S0749-596X(03)00081-0.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55*, 157–166. http://dx.doi.org/10.1016/j.jml.2006.03.007.

Vasishth, S., & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational ideas – Part I. *Language and Linguistics Compass, 10*, 349–369. http://dx.doi.org/10.1111/lnc3.12201.

Wagers, M. W. (2008). *The structure of memory meets memory for structure in linguistic cognition* (Doctoral dissertation). College Park: University of Maryland. Retrieved from <http://drum.lib.umd.edu/handle/1903/8580>.

Wagers, M. W. (2014). Syntax in forward and in reverse: Form, memory, and language processing. In A. Carnie, Y. Sato, & D. Siddiqi (Eds.), *The Routledge handbook of syntax* (pp. 409–425). New York, NY: Routledge.

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. http://dx.doi.org/10.1080/17470218.2013.858363.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85*, 79–112. http://dx.doi.org/10.1017/CBO9781139035309.009.

Warren, T., & Gibson, E. (2005). Effects of NP type in reading cleft sentences in English. *Language and Cognitive Processes, 20*, 751–767. http://dx.doi.org/10.1080/01690960500051055.

Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition, 111*, 132–137. http://dx.doi.org/10.1016/j.cognition.2008.12.011.