RESEARCH

Does antecedent complexity affect ellipsis processing? An empirical investigation

Dario Paape, Bruno Nicenboim and Shravan Vasishth
University of Potsdam, Karl-Liebknecht-Straße 24–25, D-14476 Potsdam, DE
Corresponding author: Dario Paape (paape@uni-potsdam.de)

In two self-paced reading experiments, we investigated the effect of changes in antecedent complexity on processing times for ellipsis. Pointer- or "sharing"-based approaches to ellipsis processing (Frazier & Clifton 2001, 2005; Martin & McElree 2008) predict no effect of antecedent complexity on reading times at the ellipsis site while other accounts predict increased antecedent complexity to either slow down processing (Murphy 1985) or to speed it up (Hofmeister 2011). Experiment 1 manipulated antecedent complexity and elision, yielding evidence against a speedup at the ellipsis site and in favor of a null effect. In order to investigate possible superficial processing on part of participants, Experiment 2 manipulated the amount of attention required to correctly respond to end-of-sentence comprehension probes, yielding evidence against a complexity-induced slowdown at the ellipsis site. Overall, our results are compatible with pointer-based approaches while casting doubt on the notion that changes antecedent complexity lead to measurable differences in ellipsis processing speed.

Keywords: antecedent complexity; ellipsis processing; memory pointer; self-paced reading; Bayes factor

1 Introduction

Murphy (1985) observed elevated whole-sentence reading times for the second clause in (1b) as compared to (1a).

(1) a. Jimmy swept the floor. Later, his uncle did too.
     b. Jimmy swept the tile floor behind the chairs free of hair and cigarettes. Later, his uncle did too.

In these examples, the sentence *Later, his uncle did too* contains a verb-phrase ellipsis, such that the auxiliary *did* is taken to carry the meaning of the entire verb phrase of the preceding clause. Murphy explains his experimental findings by assuming a process that copies the antecedent string into the ellipsis site. The assumption that it should take more time to transfer larger amounts of information is rather straightforward if one assumes a constant rate of throughput. Since the copied antecedent meaning is more complex in (1b) than in (1a), it is not surprising that processing time for the ellipsis should increase, given that the predication made of Jimmy's uncle becomes more complex as well.

Clearly, ellipsis is an anaphoric device, and thus superficially similar to pronouns like *he* or *she*. It can thus be assumed that some sort of memory retrieval is initiated when the ellipsis site is encountered. However, Murphy's invocation of a copying process implies that information is duplicated, unlike in the case of pronouns, which simply refer back

to an existing discourse entity.[1] Indeed, the uncle's sweeping in (1) is not identical with Jimmy's own sweeping, but refers to an independent event taking place at a different point in time.

Later studies have found no antecedent complexity effects on ellipsis processing. Using self-paced reading and a speed-accuracy trade-off (SAT) procedure, respectively, both Frazier & Clifton (2000) and Martin & McElree (2008) failed to find any evidence of longer antecedents leading to slowed processing at an ellipsis site. Based on their earlier findings, Frazier & Clifton (2001) conclude that copying is "cost-free", that is, it involves no measurable computational effort. Martin & McElree (2008) propose to do away with the copying metaphor and instead think of ellipsis as a pointer into memory. The reasoning behind the latter approach is that it is sufficient to create a link between an existing representation of the antecedent and the ellipsis site, much like creating a shortcut to a computer file, rather than creating a duplicate. This view is equivalent to what Frazier & Clifton (2005) call "structure sharing": in essence, one and the same phrase is attached in two places at once. Under this view, ellipsis is no different from pronouns in that it simply refers back to an existing linguistic entity.

In fact, having failed to find an antecedent complexity effect in a second experiment where a sentence intervened between ellipsis and antecedent, Murphy (1985) also introduces the concept of a memory pointer. He argues that comprehenders have both a structure-based and discourse-based mechanism for recovering an antecedent at their disposal. The latter is conceived of as a memory pointer and thus not subject to complexity effects (Murphy 1985: 293) while the former is argued to involve word-by-word copying. Having a clause intervene between ellipsis and antecedent arguably forces readers to fall back on the discourse-based pointer mechanism, presumably because increased distance makes the syntactic or semantic representation of the antecedent unrecoverable.

The pointer/sharing approach runs contrary to the point made above about the independence of ellipsis and antecedent. It involves, to use a term from programming, a "shallow copy" of the antecedent: the ellipsis site is interpreted by looking up the stored value from memory, but does not contain any information besides the pointer. A "shallow copy" is also used in Frazier & Clifton's (2001) account, thus rendering it equivalent in terms of predictions to the pointer/sharing view. Murphy (1985), on the other hand, assumes a "deep copy", where the information is present in both positions, as the basis for interpretation for the ellipsis. This latter conception is also often implicitly assumed in theoretical linguistics, especially if the ellipsis site is assumed to contain syntactic information (e.g., Williams 1977; Merchant 2001).[2] As Martin & McElree (2008: 882) explicitly assume that the antecedent's memory representation is accessed based on its "required morpho-syntactic, semantic, referential, and pragmatic properties", we will not subscribe to or compare any accounts which claim that ellipsis processing is exclusively syntax-, semantics- or discourse-based. In fact, the question is orthogonal to the issue of antecedent complexity, as an increase in complexity on any of the aforementioned levels will usually be accompanied by increased complexity on the other levels. However, it does strike us as being most likely that the sentence processor makes use of as much

---

[1] Note that the idea that pronouns are, underlyingly, copies of their antecedents, used to be fairly common in theoretical linguistics, but has long since become marginalized (Hankamer & Sag 1976). Pronouns cannot simply be replaced by their antecedent noun phrases in most contexts, as shown by sentences like *The caterpillar$_i$ will be able to fly when it$_i$/?the caterpillar$_i$ is a butterfly* (Grinder & Postal 1971: 269).

[2] See e.g., Dalrymple et al. (1991) for examples of empirical problems caused by a purely syntactic view.

information as it can, irrespective of the source, in order to successfully complete the retrieval.[3]

Both Murphy's (1985) experiment and the studies of Frazier & Clifton (2000) and Martin & McElree (2008) have been criticized by other scholars. Tanenhaus & Carlson (1990) note that Murphy's long antecedents may have contained temporary attachment ambiguities, while Phillips & Parker (2014) point out methodological flaws in the two more recent studies (see also Paape 2016: 3), which are discussed below.

Frazier & Clifton (2000) tested sentences like (2) in a self-paced reading paradigm.

(2)   a.   Sarah left her boyfriend last May. Tina did too.
        b.   Sarah got up the courage to leave her boyfriend last May. Tina did too.

There were only twelve items in this study, six of which were accompanied by comprehension questions that did not target the interpretation of the elliptical clause. In addition to the sentences shown above, there were two additional versions of the complex – that is, (2b) – variant of each item in which the two clauses were connected by the conjunction *and*, which means that each subject tested contributed four data points for every cell of the design. No significant effects of antecedent complexity on reading times for the elliptical clause were found, but there was a trend of 50 ms (SE: 28 ms) towards the segment *Tina did too* being read more slowly when the antecedent was complex.

As pointed out by Phillips & Parker (2014: 91), this result raises at least three major concerns. First and foremost, even though sixty subjects participated in the experiment, there could be a loss of power due to the relatively small number of observations from each participant. Moreover, and this applies to the studies of Murphy (1985) and Martin & McElree (2008) as well, measuring at the end of a sentence may introduce confounds from so-called wrap-up effects (Just & Carpenter 1980). The basic observation is that readers generally spend more time reading sentence-final regions, as well as triggering more and longer saccades in eye-tracking, which has been attributed to an overhead of comprehension processes that were not carried out during the reading of the stimulus. Whatever the exact nature of these processes, it is conceivable that their application may mask an effect of antecedent complexity on ellipsis processing. A related criticism is connected to spillover effects. Especially in self-paced reading, the effect of an experimental manipulation often only appears one or two regions downstream from where it would be expected, indicating that subjects do not finish processing each presentation region before continuing to the next one. It is thus possible that readers were still busy integrating the antecedent into the first clause when they encountered the second one, and that any observed effect of complexity is due to processing spillover.

The final concern is about the effect of task demands on reader behavior. Studies have repeatedly shown that readers adapt to experimental demands: they may fail to carry out processing steps necessary for reference assignment (Foertsch & Gernsbacher 1994), underspecify syntactic attachments (Swets et al. 2008) and leave quantifier scope ambiguities unresolved (Dwivedi 2013) in the absence of explicit motivation in the form of well-designed comprehension tests. As reported above, Frazier & Clifton (2000) did not query which meaning their readers had derived for the elliptical clause, and in fact did not ask any comprehension questions in half of the experimental trials. Martin & McElree (2008)

---

[3] An anonymous reviewer notes that a recent study by Cai et al. (2013) showed no evidence that ellipsis leads to structural priming, which can be taken to imply than syntactic information is not relevant. However, Xiang et al. (2014) did find a structural priming effect, and indeed the mere retrieval of a syntactic representation of the antecedent does not necessarily imply that any priming should be expected, given that – arguably – no structure is actively computed.

used a speed-accuracy trade-off paradigm which involved end-of-sentence grammaticality judgments. These could, however, be made correctly by simply monitoring the animacy of the unelided subject of the VP ellipsis, a strategy which does not require any deep processing of the elided part of the clause (cf. Phillips & Parker 2014: 91).

Given the concerns raised above, we feel that the issue of antecedent complexity effects in ellipsis processing has not yet been satisfactorily resolved. In our own studies reported below, we attempt to address the problems noted by the aforementioned critics. Experiment 1 resolves the issue of end-of-sentence measurements, in addition to using a non-elliptical control condition, while Experiment 2 directly tests for an influence of task demands on antecedent complexity effects. First, however, yet another perspective on the possible effects of antecedent complexity on ellipsis processing will be introduced; it predicts that instead of slowing down the interpretation process, increased complexity of the antecedent should result in a speedup.

Hofmeister (2011) investigated the processing of cleft sentences, which contain a filler-gap dependency between the clefted constituent and the position it was extracted from. In (3), the phrase *a (...) communist* is the object of the verb *banned*, and thus has to be retrieved from memory when the verb is read to compute the meaning of the clause.

(3)   It was [a communist]/[an alleged communist]/[an alleged Venezuelan communist] who the members of the club banned from ever entering the premises.

In a self-paced reading study, Hofmeister found that reading times right after the verb *banned* decreased with the complexity of the filler phrase. Further experiments showed that increasing the semantic specificity of the antecedent also decreased processing times at the gap when string length was kept constant (*which person* vs. *which soldier*), but that making the filler difficult to process (*the lovable military dictator*) resulted in a slowdown rather than a speedup. Hofmeister concludes that more elaborate descriptions of retrieval targets aid memory access as long as they are "typical" (*ruthless military dictator* showed an advantage over *dictator*). He proposes that features which are closely associated (*ruthless – dictator*; *wealthy – celebrity*) will speed up access to the memory target because activation spreads from feature to feature.

Coming back to ellipsis, if the event description encoded by the retrieval target – that is, the verb phrase in (1) – becomes more elaborate, it should become easier to access. Informally, when a reader of (1b) encounters the word *did*, remembering that Jimmy did something involving hair and cigarettes might facilitate access to the sweeping event described by the antecedent. While this is precisely the opposite of what Murphy (1985) observed, it is possible that any advantage due to elaboration was lost due to idiosyncrasies of the items used in his study. In Murphy's (1a), it does not matter whether the floor was still dirty when Jimmy's uncle swept it, while in (1b) it clearly was, which requires a laborious inference on part of the reader.

In the two self-paced reading studies we present in this paper, we investigated the processing of ellipses with antecedents of varying complexity. In order to broaden the scope of the inquiry, Experiment 1 focused on German instead of English. Since VP ellipsis does not exist in German, stimulus sentences in this experiment contained a construction known as bare argument ellipsis, also called "stripping". Experiment 1 improved upon previous studies that did not feature control conditions without ellipsis – Martin & McElree (2008) being a notable exception – and featured a subset of comprehension questions that directly targeted the interpretation of the elliptical clause. Experiment 2 addressed the concern originally raised by Phillips & Parker (2014) that superficial processing may have played a role in the studies of Frazier & Clifton (2000) and Martin & McElree (2008). To this aim, we manipulated the types of comprehension questions that participants had to

answer, much like Swets et al. (2008) did when investigating the resolution of temporary syntactic ambiguity.

2 Experiment 1

Bare argument ellipsis or 'stripping' deletes an entire clause, with the exception of one constituent, plus an adverb in some cases (Hankamer & Sag 1976). A German example is given in (4), where the second of the conjoined clauses is understood to mean *John wanted to jump over the fence as well.*

(4)   *German*
    Peter wollte über den Zaun springen und Johann ebenfalls.
    Peter wanted over the fence jump and John as.well
    'Peter wanted to jump over the fence and John (did) as well.'

Stripping targets constituents which are larger than VP, as evidenced by the fact that the modal is deleted along with the lexical verb. Apart from this, we know of no reason why the processing of stripping constructions should differ fundamentally from that of VP ellipsis in English, other than that cues for a different kind of retrieval target are set, and that the cuing element in this case is an adverb rather than an auxiliary. As with VP ellipsis, when the gap site is identified, the processor needs to look for a suitable antecedent whose meaning (or structure) the gap is to be identified with. In this example, the antecedent consists of the string *wollte* über *den Zaun springen*, 'wanted to jump over the fence'.

2.1 Materials

A sample stimulus from Experiment 1 is shown in (5). Diamonds indicate the boundaries of presentation regions during the experiment. The study employed a 2 × 2 design with the experimental factors antecedent complexity (simple vs. complex) and elision (ellipsis vs. control). A total of twenty-eight items were created. The stimuli are listed in Appendix A. Ninety filler items featuring a variety of constructions were also presented during each experimental session.

(5)   **Simple antecedent**
    Die Armee ◊ räumte ◊ einige wichtige Feldlager ◊ . . .
    The army cleared some important field camps
    'The army cleared some important field camps ...'

    Complex antecedent
    Die Armee ◊ ließ ◊ nach dem Gefecht ◊ einige wichtige Feldlager
    The army let after the battle some important field camps
    ◊ räumen ◊ . . .
    clear
    'The army had some important field camps cleared after the battle ...'

    Continuation
    . . . und ◊ der kluge Befehlshaber ◊ der Aufständischen ◊ . . .
    and the clever commander of the insurgents
    '... and the clever commander of the insurgents ...'

    Ellipsis
    . . . ebenfalls ◊ . . .
    as.well

Control

... rückte vor   ◊ ...
    advanced

_____

Continuation

... ohne   ◊ allerdings ◊ die Sicherung   ◊ seiner Versorgungslinien ◊ zu
    without   however   the protection   of his supply lines   to
vernachlässigen.
neglect
'... without, however, neglecting the protection of his supply lines.'

All experimental sentences featured the same structure, namely an antecedent clause connected to another clause via the conjunction *und*, 'and'. The critical region is the final word of the second clause, which is either the adverb *ebenfalls*, 'as well', or an intransitive lexical verb. The adverb signaling the ellipsis remained the same across all items while the verbs in the control conditions differed. Antecedent complexity was manipulated by adding a modal verb or auxiliary and an adjunct to the simple version of the first clause.[4] The sentence continues after the critical region in order to prevent wrap-up effects due to periods and allow for spillover.

2.2 Participants

Sixty native speakers of German participated in the experiment. These were recruited from the Vasishth Lab's subject pool at the University of Potsdam, which is administrated and maintained through ORSEE (Greiner 2015). Each subject was either paid 6 € or received course credit. Informed consent from the participant was obtained before each experimental session. The experiment complied with the June 1964 Declaration of Helsinki (carried out by the World Medical Association and entitled "Ethical Principles for Medical Research Involving Human Subjects"), as last revised. In accordance with German NSF (DFG) guidelines, for experiments with unimpaired adult populations, the ethics approval is required by the Principal Investigator (in this case, Prof. Dr. Shravan Vasishth).

2.3 Procedure

The experimental stimuli were presented in a latin-square design using the Linger software written by Douglas Rohde (Rohde 2003), along with the filler items. Presentation order was randomized at runtime. Participants were instructed to read silently at their normal pace. Each trial started with a white screen that was displayed for 1000 ms and that could not be skipped. The sentence was then shown in masked form, that is, with all characters except spaces replaced by underscores (_). Participants pressed the space bar to replace the underscores with the corresponding regions of the sentence, displayed in 20 pt Courier New font. Presentation was non-cumulative, that is, previous regions reverted back to underscores upon continuation. Times between button presses were recorded. After every sentence, a statement was shown that participants were required to judge as being either true or false, based exclusively on the information given by the stimulus. For instance, a subject reading the simple/ellipsis version of (5) would have been required to judge the statement *A clever commander had some important field camps cleared* (true) while a subject reading the complex/control version would have judged

_____

[4] The causative verb *lassen*, 'let', is, strictly speaking, neither a modal verb nor an auxiliary, even though it is known to behave like one of the former in most respects (Bader 2014). For the purposes of the present study, the central requirement was that *lassen* embeds an infinitive and is deleted along with it in ellipsis. Other items featured more typical modals such as *wollen*, 'want', *müssen*, 'must', and *sollen*, 'should', or the auxiliary *haben*, 'have'.

the statement *A clever commander had to clear some important field camps* (false). The ratio of true to false statements was 1:1 across the entire experiment. Out of fifty-six possible cases (twenty-eight items times two conditions), twenty-one comprehension tests targeted the interpretation of the ellipsis. Other statements targeted either the antecedent or other parts of the stimulus sentences. Participants were given the opportunity to take a break after completing half of the experiment.

2.4 Predictions

If ellipsis is interpreted via a memory pointer mechanism (Frazier & Clifton 2005; Martin & McElree 2008) or, equivalently, a cost-free whole-clause copying mechanism (Frazier & Clifton 2001), we expect no effect of the antecedent complexity at the critical region – that is, the ellipsis site – in the elided conditions. However, under the copying account of Murphy (1985), we expect longer reading times at the critical region for sentences with complex antecedents in the ellipsis conditions only. As no clause intervenes between antecedent and ellipsis site, Murphy's theory predicts that readers should not fall back on a discourse-based processing mechanism, which would otherwise lead us to expect no effect of antecedent complexity. Finally, if more elaborate antecedents are easier to retrieve from memory, as would be expected given the findings of Hofmeister (2011), reading times at the critical region should be shorter for sentences with complex antecedents in the ellipsis conditions.

Note that both the Murphy (1985) and Hofmeister (2011) accounts predict an interaction between antecedent complexity and elision. This is important because antecedent complexity is completely confounded with the ellipsis site's position in the sentence. Any main effect of the complexity manipulation could thus be due to changes in participants' reading speed as they progress through the sentence (Ferreira & Henderson 1993; Demberg & Keller 2008). Martin & McElree (2008) circumvented this problem by adding material between antecedent and ellipsis in the simple antecedent conditions, which, however, increases the distance between the end of the antecedent clause and the ellipsis site, as well as introducing the possibility that the processing of the additional information may interfere with the encoding or retrieval of the antecedent. Being faced with two less-than-optimal alternatives, we opted to stay as close as possible to the designs of Frazier & Clifton (2000) and Murphy (1985), which did not keep sentence length constant across conditions.

Looking more closely at the results of Martin & McElree (2008), it should be noted that according to Foraker & McElree (2011), the failure to find an effect of a manipulation on processing speed in an SAT paradigm by itself does not entail that there should also be no effect on reading times in comparable self-paced reading or eye-tracking studies. Foraker & McElree (2011) argue that even if only the asymptotic accuracy – the highest level of accuracy that participants are able to reach with their grammaticality judgments – is affected in SAT, reading times in self-paced reading or eye-tracking may differ between conditions due to retrieval failures or low-quality interpretations. More specifically, a drop in asymptotic accuracy in SAT may translate to higher reading times due to reprocessing (McElree & Nordlie 1999). Martin & McElree (2008) largely failed to find effects of antecedent complexity on asymptotic accuracy, with the exception of their Experiment 6, where an additional full noun phrase within the antecedent lowered accuracy. Based on this isolated result, higher reading times should be predicted for antecedents containing more full noun phrases. However, in Martin & McElree's other experiments, which also included an eye-tracking study, the presence of additional noun phrases in complex antecedents did not measurably affect accuracy or reading times, calling the result of Experiment 6 into question. We thus take Martin & McElree's evidence to point more

strongly in the direction of there being no effect of antecedent complexity on ellipsis processing across paradigms, and indeed this appears to be the position adopted by the authors.

2.5 Data analysis

All data from the first participant were discarded before analysis as this session was considered a trial run, which revealed several minor mistakes. The remaining data were analyzed using the statistics software R (R Core Team 2015). Linear mixed-effect models were fit to reading times and question response accuracies with the package rstanarm (Gabry & Goodrich 2016), which provides an interface between R and the Stan programming language for Bayesian statistical inference (Stan Development Team 2016). The data and code for both experiments will be released with the publication of this article.

One advantage of Bayesian inference in Stan is that a hierarchical linear model can almost always be fit with full variance-covariance matrices for subject and item random effects (Sorensen et al. 2016); this is often difficult to achieve with the lme4 function (Bates et al. 2015b; see Bates et al. 2015a for further discussion). Another advantage is the more straightforward interpretation of results in a Bayesian setting. Instead of computing confidence intervals, which somewhat unintuitively refer to hypothetical repeated sampling (Hoekstra et al. 2014), a Bayesian credible interval specifies plausible values of the parameters given the data at hand. This makes inference much more straightforward compared to Null Hypothesis Significance Testing (see Nicenboim & Vasishth 2016 for a review).

Reading times below 150 ms, which are (arguably) unlikely to be generated by linguistic processes, were removed prior to analysis; this resulted in a loss of less than 1% of data. The experimental factors were sum-coded. For the factor antecedent complexity, the complex conditions were coded as 1 and the simple conditions were coded as –1, respectively. For the factor elision, the ellipsis conditions were coded as 1 and the control conditions were coded as –1. As visual inspection of the reading time distributions suggested some amount of heteroscedasticity in the data, the Box-Cox procedure (Box & Cox 1964) was applied, which suggested reciprocal transformation of reading times (1/RT) and logarithmic transformation of question-response times. Reciprocal reading times were multiplied by a factor of –1000 to allow for easier interpretation. All models included random intercepts and slopes by subjects as well as by items for each estimated parameter, including interaction parameters. The prior distribution for each estimated parameter was a normal distribution with mean zero and a standard deviation of 2.5, except for the intercept, for which a standard deviation of 10 was used. The LKJ prior (Lewandowski et al. 2009) with parameter 1 was used for the variance-covariance matrices of the random effects for subjects and items;[5] this imposes a regularization on the prior distribution of the variance-covariance matrix (see Stan Development Team 2016 for details, and Sorensen et al. 2016 for a tutorial intended for psycholinguists). Besides fitting models to individual regions of interest, as is commonly done in psycholinguistics, we also fitted a model that took into consideration all data points from the second-to-last region leading up to the ellipsis site (crit–2) to the second region after the ellipsis site (crit + 2). The region predictor was coded using a successive differences contrast, meaning that the model is estimating the *differences* in processing times between each two adjacent regions, starting with region crit–2. The region-by-region analyses can thus be seen as nested comparisons within the overall model (see

---

[5] An LKJ prior with parameter 1 assigns equal probability mass to all possible correlation values. An *sd*(0,10) prior on the intercept gives values between ±19.6 on the transformed –1000/RT scale, and values close to ±1600 ms on the original scale. This uninformative prior serves as a sanity check to see if the model can recover a sensible intercept.

Nicenboim et al. 2016). To account for the fact that reading times within the same trial are not independent, we added a random intercept by trial to the model.

Four sampling chains with 4000 iterations each were run for each model, with a warm-up period of 2000 iterations. We report the estimated parameters, along with their 95% credible intervals and the posterior probability that the parameter's true value is greater than zero. We judge there to be evidence for an effect if zero is not included in the associated 95% interval. We consider there to be weak evidence for an effect – which is to be distinguished from the effect itself being weak – if zero is within the interval but the probability of the parameter being above or below zero is high (>95%).

2.6 Results

2.6.1 Question responses

Question response accuracy was 88% for all items and 85% for target items. The analysis of response accuracies revealed no effects of the experimental manipulations. However, there is some evidence of response times being elevated for ellipsis versus control sentences ($\hat{\beta}$ = 0.031, CrI: [–0.004, 0.067], Pr ($\hat{\beta}$ > 0) = 0.96). See Figure 2 for the results.

2.6.2 Reading times

Figure 1 plots back-transformed mean reciprocal reading times (see above) by region, along with 95% confidence intervals. Figure 3 shows the results for individual regions of interest. Analyzing the regions between antecedent and ellipsis site is critically important because of the concern that processing spillover from the antecedent may influence reading times at the critical regions. In region crit–2, sentences with complex antecedents were read faster than sentences with simple antecedents ($\hat{\beta}$ = –0.029, CrI: [–0.056, –0.002], Pr ($\hat{\beta}$ > 0) = 0.02), indicating that there is no spillover from processing the antecedent. The complexity advantage – which may reflect a speedup due to having processed additional words – persists into region crit–1 ($\hat{\beta}$ = –0.025, CrI: [–0.049, 0.001], Pr ($\hat{\beta}$ > 0) = 0.03), but disappears as soon as the critical region is encountered. If anything, antecedent complexity *increased* reading times at the critical region, both for
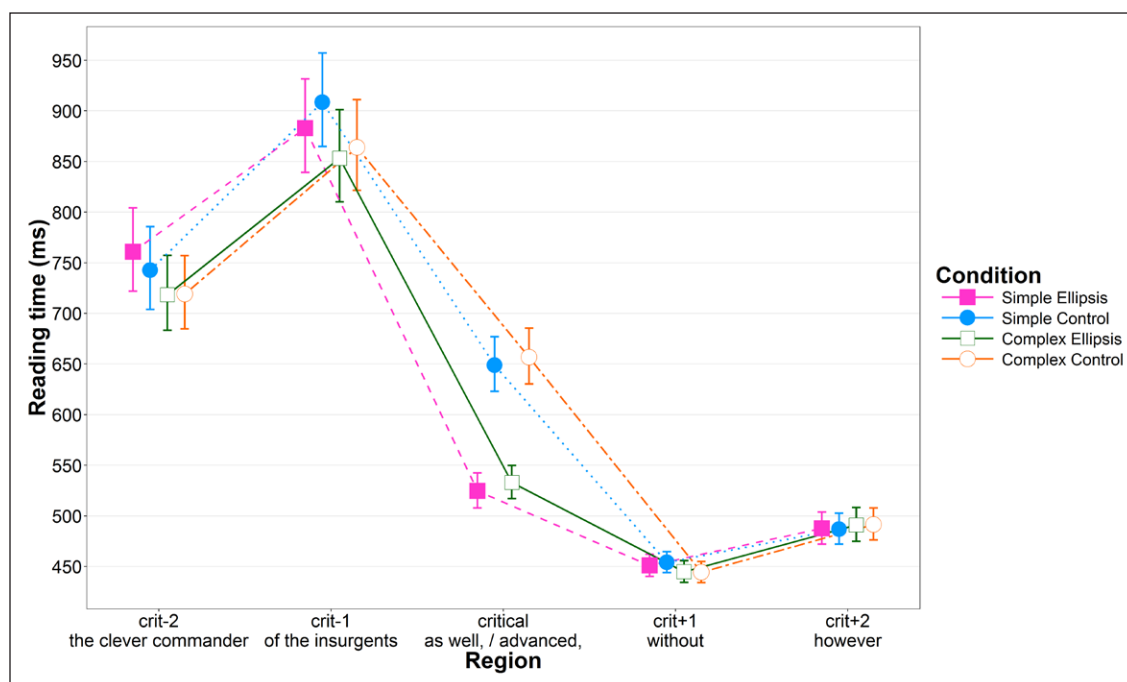
Figure 1: Experiment 1 – Mean reading times (back-transformed, on reciprocal scale) by region and condition. Error bars represent 95% confidence intervals.
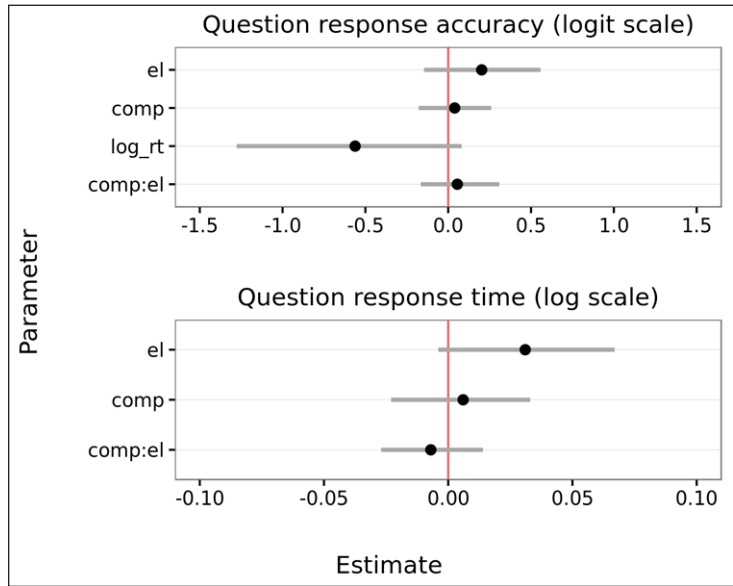
Figure 2: Experiment 1 – Question response times and accuracies: caterpillar plot of means and 95% credible intervals for parameters of interest. el = elision, comp = complexity, log_rt = log response time.
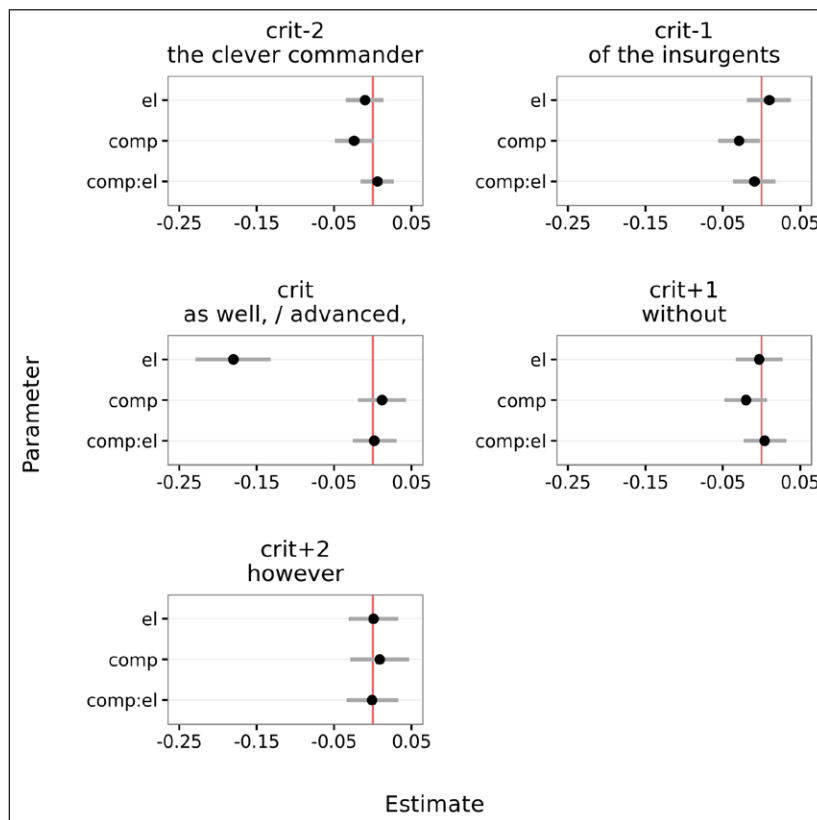
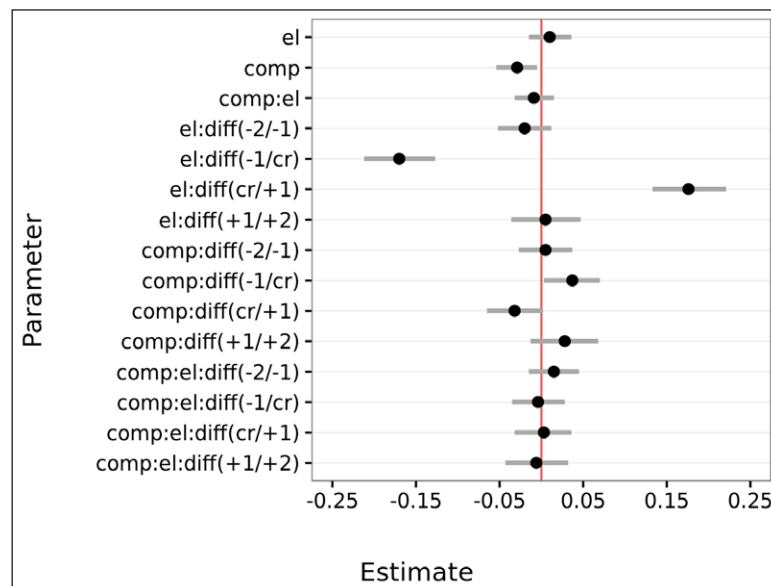

Figure 3: Experiment 1 – Caterpillar plot of means and 95% credible intervals for parameters of interest, separate models fit across regions crit–2 through crit+2. el = elision, comp = complexity.

ellipsis and control sentences, but the evidence is very weak ($\hat{\beta} = 0.012$, CrI: [–0.019, 0.043], Pr ($\hat{\beta} > 0$) = 0.78). The critical region shows a main effect of elision, such that ellipsis is processed faster than the lexical verbs in the control conditions ($\hat{\beta} = -0.180$, CrI: [–0.229, –0.132], Pr ($\hat{\beta} > 0$) ≈ 0). In the region following the ellipsis (crit + 1), there is

again weak evidence for a complexity-induced speedup ($\hat{\beta}$ = –0.020, CrI: [–0.048, 0.007], Pr ($\hat{\beta} > 0$) = 0.08). There was no evidence for an interaction between antecedent complexity and elision in any region.

Figure 4 shows the estimated parameters from the model fitted to regions crit–2 through crit + 2 together. In addition to an overall complexity advantage ($\hat{\beta}$ = –0.029, CrI: [–0.054, –0.005], Pr ($\hat{\beta} > 0$) = 0.01), sentences with complex antecedents show a smaller difference between region crit–1 and the critical region ($\hat{\beta}$ = 0.037, CrI: [0.003, 0.070], Pr ($\hat{\beta} > 0$) = 0.98), but a larger difference between the critical region and region crit + 1 ($\hat{\beta}$ = –0.032, CrI: [–0.065, 0.002], Pr ($\hat{\beta} > 0$) = 0.03).[6] This finding matches the patterns observed in the by-region analyses. Elision, meanwhile, increases the difference in processing times between regions crit–1 and the critical region ($\hat{\beta}$ = –0.170, CrI: [–0.212, –0.127], Pr ($\hat{\beta} > 0$) $\approx$ 0), but decreases the difference between the critical region and region crit + 1 ($\hat{\beta}$ = 0.176, CrI: [0.133, 0.221], Pr ($\hat{\beta} > 0$) $\approx$ 1), matching the main effect of elision observed in the critical region itself.

While there was no evidence of an interaction between the elision and complexity manipulations at the critical region, the main effect of complexity that is visible in the overall analysis is of theoretical interest. As we were interested in further investigating the effect of antecedent complexity on reading times for the critical region, we subjected the relevant coefficient from the single-region model, whose estimate showed only very weak evidence for being positive, to a Bayes Factor analysis. A hypothesis test based on the Bayes Factor provides a way to quantify the support for the model under which the observed data are most likely (Wagenmakers et al. 2010). We chose to perform multiple order-restricted analyses, meaning that sampling was restricted to either only positive or only negative coefficient values, respectively, in order to better gauge the amount of evidence for or against the coefficient in question being different from zero in a given direction. By using left- or right-truncated prior distributions,


Figure 4: Experiment 1 – Caterpillar plot of means and 95% credible intervals for parameters of interest, one model fit across regions crit–2 through crit+2. *diff* parameters represent successive differences between regions, in linear order. el = elision, comp = complexity.

---

[6] Both difference parameters are negative, which means that an interaction with a negative sign indicates a larger difference while one with a positive sign indicates a smaller difference.

it is possible to quantify support against the null hypothesis and also in favor of it, relative to a directed alternative hypothesis. Additionally, the conclusions from the Bayes Factor test are more conservative than those based on credible intervals: Even a 95% credible interval that does not include zero does not guarantee a high value of the Bayes Factor, that is, it does not guarantee strong support for the alternative hypothesis (Wagenmakers et al. 2010).

We used the Savage–Dickey density ratio method (Dickey & Lientz 1970) to compute the Bayes Factor, following Wagenmakers et al. (2010). Even though the posterior distributions for the model parameters are generally not sensitive to the prior settings, the Bayes Factor is acutely so. When priors are too wide (too uninformative), the alternative hypothesis assigns too much prior mass to values that yield very low likelihoods. This in turn means that without proper specification of priors, the null hypothesis would be always more likely than the alternative hypothesis, since its prior mass in concentrated in zero. Three normal distributions of different widths were used as priors on the complexity coefficient (on the transformed scale): N (0, 0.05), N (0, 0.025) and N (0, 0.005). Model specifications were otherwise left unchanged. With N (0, 0.05), 95% of the probability mass will be in the interval [0, 0.1] for a left-truncated normal distribution and in the interval [–0.1, 0] for a right-truncated normal distribution. The corresponding values on the original scale are 69 ms, 34 ms and 7 ms, respectively. Figure 5 shows the
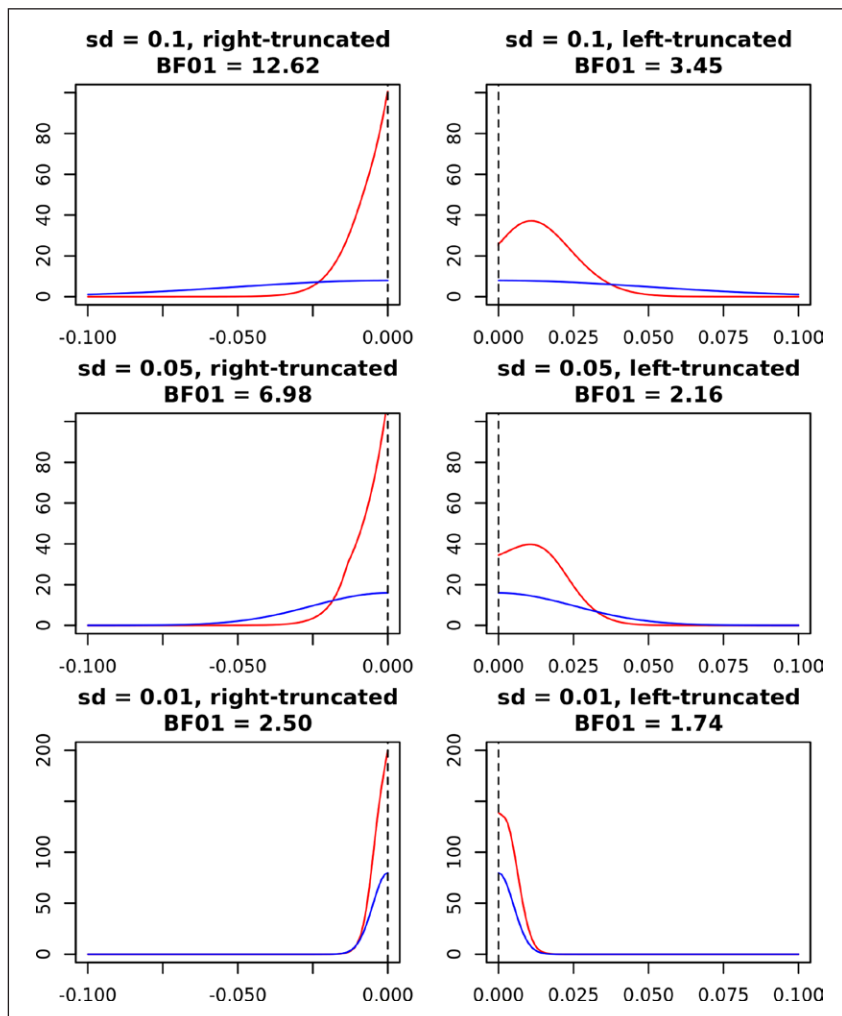
Figure 5: Experiment 1 – Density plots for the Bayes factor analysis. Blue = prior density, red = posterior sample density.

calculated Bayes factor values for the different priors, along with density plots for the prior distributions versus posterior samples.[7]

For the right-truncated priors, the value of the Bayes factor depends heavily on the spread of the distribution: For the widest prior, the null hypothesis is more than twelve times as likely as the alternative hypothesis that the complexity effect is negative, while for the narrowest prior it is still between two and three times as likely. For the left-truncated priors, the null hypothesis is between two and three-and-a-half times as likely to be true, depending on the spread – and thus the informativeness – of the prior. Note that unlike for left-truncated priors, the point of maximum probability density for the posterior samples given right-truncated priors is always at zero. On the whole, the analysis shows that for all but the most narrow distributions the prior restriction that the complexity effect should be negative or null leads to more evidence in favor of the null hypothesis compared to when the effect is restricted to be positive or null. There is thus evidence that the effect is probably not negative, and more likely to be null than positive, even though the latter conclusion is only weakly supported if one adheres to common interpretation guidelines for the Bayes factor (Raftery 1995).

2.7 Discussion

Three main results were obtained in the current study:

(I)　　　Ellipsis was processed faster than the lexical verbs used in the control conditions.

(II)　　Overall, having processed a longer and more complex antecedent led to faster reading times across later regions.

(III)　　At the critical region, the speedup was temporarily suspended. An analysis based on the Bayes factor yielded evidence in favor of a null effect of the complexity manipulation at the critical region.

Finding (I) may be trivially explained by the fact that the critical region was shorter in the ellipsis compared to the control conditions for most items. The prediction of an interaction between the elision manipulation and antecedent complexity was not borne out in the data, a result that is most consistent with pointer-based accounts of ellipsis resolution (Frazier & Clifton 2001, 2005; Martin & McElree 2008). Furthermore, finding (III) suggests that the overall speedup induced by the complexity manipulation was nullified at the critical region, rather than turning into a slowdown, further supporting pointer-based approaches. Even if there had been evidence of such a slowdown – which, given the Bayes factor results, would need to be of a very small magnitude – we would still have needed to explain why it would appear in both the ellipsis and control conditions (see discussion below).

We believe that finding (II) has a mechanistic explanation: readers tend to read faster if they are deeper into the sentence already. The "complexity"-induced speedup would thus be a length or, equivalently, a linear position effect. Given our initial predictions, we found no evidence that increased antecedent complexity slowed down the processing of the ellipsis, as predicted by Murphy's (1985) account, or contrariwise led to speedier processing of the ellipsis, as predicted by the account of Hofmeister (2011). Rather, the

---

[7] Note that as we are using the logspline package (Kooperberg 2016) in R to estimate the marginal posterior density at zero, there is some variation in the calculated Bayes factors when the computations are run multiple times. In our test runs, each Bayes factor varied within the limits of approximately ± 0.5 of the values shown in Figure 5.

Bayes factor analysis showed that the data are largely inconsistent with the assumption that increased complexity leads to faster processing.

2.8 A potential influence of within-sentence wrap-up

The results at the critical region warrant closer inspection, as one might argue that the observed temporary suspension of the speedup effect could be due to wrap-up caused by the comma at the end of the second conjunct. If wrap-up reflects integration processes at the end of a clause, since integrating more complex meanings takes longer, readers will possibly spend more time on the final region of the second conjunct if the first conjunct contains more information. This may then momentarily cancel out the speedup that is visible before and after the critical region. When designing the experiment, we avoided having the ellipsis followed by a period, neglecting that commas also create wrap-up effects, albeit of a smaller magnitude (Warren et al. 2009). In our defense, it is quite impossible to study clausal ellipsis without having the end of the elided clause marked somehow in the input. In any case, there is evidence from eye-tracking suggesting that wrap-up at punctuation marks such as commas is not influenced by the complexity of the sentence (Rayner et al. 2000; Warren et al. 2009), which casts doubt on the assumption the complexity effect observed at the critical region is only due to the comma.

If anything, one would need to claim that the position-based speedup in reading that has been observed repeatedly (Ferreira & Henderson 1993; Demberg & Keller 2008) is completely suspended during wrap-up. As a quick check of this assumption, we fitted a Bayesian linear mixed-effects model to the data from our filler items. In this model, linear position of the presentation region within the sentence and the presence or absence of a comma were used as predictors. The comma factor was sum-coded with comma present being coded as 1, and region position was entered as a continuous predictor. The model revealed that there was indeed a position-related speedup ($\hat{\beta} = -0.010$, CrI: [-0.021, 0.000], Pr ($\hat{\beta} > 0$) = 0.027), as well as a comma-induced slowdown ($\hat{\beta} = 0.057$, CrI: [0.025, 0.091], Pr ($\hat{\beta} > 0$) ≈ 1), and an interaction with a negative sign: the speedup effect appears to be *stronger* rather than weaker when a comma is present ($\hat{\beta} = -0.008$, CrI: [-0.017, 0.001], Pr ($\hat{\beta} > 0$) = 0.038). This implies that the presence of a comma probably did not result in a suspension of the speedup effect observed in our experimental items.

Given these findings, the possibility arises that the speedup was still in effect at the critical region, but was counteracted by a complexity-induced slowdown in the vein of Murphy (1985), resulting in the two effects canceling each other out. Under this assumption, however, one is left asking why the slowdown should also be present in the control conditions.

2.9 A possible issue of parallelism

There may be other reasons for not expecting an effect of the manipulation in our materials. Particularly, our use of the conjunction *und*, 'and', might be critical to understanding our failure to observe an interaction between antecedent complexity and ellipsis processing. The results of a cross-modal priming study by Callahan et al. (2010) are informative in this regard. In their Experiment 2, Callahan et al. presented sentences like (6) auditorily. Words that were either related or unrelated to the verb *read* in the initial clause (related: *reviewed*, unrelated: *reserved*) appeared on the screen at the positions marked in the example. Subjects were required to read these words aloud.

(6)   The doctor read the chart of the child with the broken arm [1] during his morning rounds, and [2] the insurance agent in [3] the tacky suit did as well [4] in order to become more familiar with the case.

Results showed that naming responses to related words were faster at probe positions 3 and 4, but not at positions 1 and 2. Experiment 1 used only probe positions 1 and 2, revealing a priming effect at position 2, but not at position 1. Despite the priming effect at the conjunction itself not appearing consistently, Callahan et al. (2010) conclude that material from the first clause is reactivated during the processing of the second clause. The conjunction *and* arguably induces an expectation of parallelism, causing the retrieval and subsequent maintenance of the verb *read*, or possibly of the entire associated proposition, allowing for easier integration with the second conjunct. Callahan et al. (2010) suggest that active maintenance of antecedent information may be achieved through repeated retrievals prior to the ellipsis site which are cued by the conjunction. Even though parallelism has long since been known to facilitate the processing of coordinate structures (Frazier et al. 1984; Frazier et al. 2000), Callahan et al. (2010)'s *sustained reactivation hypothesis* is, to our knowledge, the first account to explicitly link this observation to working memory.

If the presence *und*, 'and', in our stimuli led participants to assume parallelism between the conjuncts, causing them to actively maintain information from the antecedent clause, there is an alternative explanation for the prolonged speedup effect we observed: participants were simply eager to reach the end of the second conjunct, since this is the point where the two propositions can be integrated. Crucially, sustained reactivation also obviates the need for a laborious retrieval at the critical region, since the necessary information is already available, thus predicting no detrimental effect of the complexity manipulation, apart from possible costs associated with discourse integration.

Even without sustained reactivation being a factor our stimuli, the lack of an interaction between antecedent complexity and elision can be explained if one assumes that lexical verbs can also trigger retrievals. This might be true especially in coordinate structures, where parallelism reinforces the semantic association between the conjuncts. Indeed, the control conditions in many of our sentences imply a causal connection between the two propositions, such as the commander advancing after the enemy's field camps have been cleared in (7).

(7)   The army cleared some important field camps and the clever commander of the insurgents advanced.

Pointer-based approaches (Frazier & Clifton 2001; Martin & McElree 2008) can account for the result by claiming that retrieval time is negligible across conditions, and that any complexity-induced slowdown reflects integration difficulty *after* retrieval. However, one would then need to assume that this integration difficulty is limited to *and*-conjoined sentences like the ones used in our study: Frazier & Clifton (2001) found no complexity effect for two-sentence discourses – but recall the study's limitations noted in the introduction – and Martin & McElree (2008) found no complexity-induced change in ellipsis processing times for *but*-conjoined sentences.

2.10 A more precise notion of complexity-based facilitation is needed

Assuming that retrieval takes place in both the ellipsis and control conditions, the observed processing pattern would be more in line with the reasoning of Murphy (1985), where it takes more time to copy more information from the antecedent, than with that of Hofmeister (2011), where elaboration should lead to facilitation. Indeed, our analyses showed more evidence for the former view than the latter. However, as was pointed out before, Hofmeister (2011: 395) assumes that not all kinds of elaboration aid retrieval; only strongly associated features of a memory trace are predicted to have a facilitatory effect. Uncommon feature combinations (*lovable dictator*), while increasing encoding time,

will impede retrieval instead of providing easier access to the target. While Hofmeister's results show that there is no direct connection between encoding time and retrieval time, it is by no means clear whether the elaboration provided by the complexity manipulation in the current study should have yielded any additional facilitatory anchors for memory access. The answer would depend, among other things, on whether the component parts of the antecedent are visible to the retrieval probe. If we assume that the search process that is initiated when a clausal ellipsis is encountered focuses on finding a phrase containing a verb, which is the semantic core of a clause, it might ignore any adjuncts or auxiliaries attached to it. If the search process is serial, the presence of such elements may even result in longer processing times. Taken at face value, however, the theory of Hofmeister (2011) should predict facilitation for our stimuli, given that clausal adjuncts are to sentence meaning what adjectives, as used by Hofmeister, are to a noun phrase, that is, elaborative modifiers. Thus, if the presence of an adjective influences the retrieval process, so should the presence of a clausal adjunct.

2.11 Conclusion

In short, Experiment 1 showed evidence in favor of a null effect of antecedent complexity on ellipsis processing times. The results should, however, be interpreted with a certain amount of caution. On the methodological side, one important shortcoming is that the experiment used sentences conjoined by *und,* 'and', possibly causing the control conditions not to work as intended. Our second study sidesteps the issue of parallelism by using *but-* instead of *and-*coordinated sentences. Unlike *and, but* evokes no expectation of parallelism between the two conjuncts, and indeed parallelism does not facilitate processing for *but-*conjoined sentences (Knoeferle 2014). The main goal of Experiment 2 was to investigate whether antecedent complexity effects in ellipsis processing are sensitive to task demands, as suggested by Phillips & Parker (2014) and Paape (2016). The design is inspired mainly by Swets et al. (2008)'s investigation of parsing preferences for a temporary syntactic ambiguity.

3 Experiment 2

Drawing from the literature on "good-enough" processing (e.g., Christianson et al. 2001; Ferreira 2003), Swets et al. (2008) explored whether asking different kinds of comprehension questions would influence readers' on-line processing of syntactically ambiguous sentences in a self-paced reading experiment. The construction in question involves a relative clause whose attachment site is initially not obvious, as shown in (8). The gender of the reflexive *himself/herself* disambiguates the structure towards attachment to either the first NP (N1) or the second (N2) in (8b, c), but not in (8a).

(8)   a.   The maid of the princess who scratched herself . . .   **ambiguous**
      b.   The son of the princess who scratched himself . . .   **N1 attachment**
      c.   The son of the princess who scratched herself . . .   **N2 attachment**
           . . . in public was terribly humiliated.

Subjects were divided into three groups according to the kind and frequency of comprehension questions that appeared along with the experimental sentences. One group of participants was asked questions that targeted the interpretation of the relative clause, such as *Did the maid scratch in public?* A second group answered questions that did not target the relative clause, and indeed did not require much attention to the sentences' contents, such as *Was anyone humiliated?* A third group was also asked these superficial questions, but only on one out of every twelve trials.

Swets et al. (2008) found that participants expecting questions about the relative clause attachment took longer to read the post-disambiguation region if the attachment had been disambiguated toward N1 (8b) than for both of the other conditions. The pattern for readers in the two other groups looked different: they were faster in the ambiguous condition (8a) than in both the N1 and N2 conditions. These results indicate that readers' syntactic processing strategies may change according to task demands. If participants know that their interpretation of an ambiguous sentence will be probed, they appear to preferentially choose one possibility, namely N2 attachment. If, however, participants do not have to worry about their interpretation being queried explicitly, the enjoy a processing advantage due to the possibility of not making an attachment decision at all. This is commonly referred to as underspecification.

Given that effects of task demands have also been observed in discourse processing (Foertsch & Gernsbacher 1994), it is conceivable that people have more than one strategy available for the resolution of ellipsis. Another possibility is that readers can be somewhat selective in terms of what information they retrieve – or, alternatively, maintain and integrate – at the ellipsis site. If comprehension of the elliptical clause is not probed too deeply, they might even opt to not resolve the anaphor at all. This latter view is rather extreme, given the implication that readers never make an effort to understand experimental stimuli unless explicitly motivated to do so. One might argue that since reliable effects of experimental manipulations can be observed even in studies which feature no or only shallow tests of comprehension, there must be some intrinsic motivation to interpret sentences even when there is no payoff. While this is a valid point, it is by no means clear whether we can rely on the compliance of our subjects in all cases, especially in light of recent findings on "good-enough" processing.

While Experiment 1 investigated bare argument ellipsis ("stripping") in German, Experiment 2 used English VP ellipsis constructions, much like the aforementioned studies of Frazier & Clifton (2000) and Martin & McElree (2008). As the discussion of Experiment 1 suggested, the control conditions used in the previous study may not have served their purpose as intended, so for Experiment 2 we dispensed with them. Instead, subjects were divided into two groups which received different kinds of comprehension probes during the experiment. Since many of these were directly related to the interpretation of the elided VP, we can assume that any group-specific effects we observe will be connected to the presence of ellipsis, rather than to other aspects of the stimuli.

3.1 Materials

A sample stimulus from Experiment 2, along with two of the associated comprehension probes, is shown in (6). As before, diamonds indicate the boundaries of presentation regions during the experiment. The experimental factors used in this study were antecedent complexity (simple vs. complex) and probe type (superficial vs. detailed). In the current study, simple antecedent clauses always contained only a simple object NP (see below), while in complex antecedent clauses this object NP in turn contained a genitive modifier as well as additional adjectives. Note that unlike in Experiment 1, the antecedent complexity manipulation did not change the number of presentation regions. Probe type remained constant throughout each experimental session and divided subjects into two groups. The study thus employed a 2 (within-subjects) × 2 (between-subjects) design. A total of thirty-six items and one-hundred and sixty fillers were presented in random order during each experimental session. The stimuli are listed in Appendix B.

(9)   **Antecedent preamble**
        The advanced students ◊ loved ◊ . . .

___

        Simple antecedent
        . . . the afternoon session, ◊ . . .

___

        Complex antecedent
        . . . the late afternoon session's many illustrative examples, ◊ . . .

___

        Continuation
        . . . but ◊ as of late ◊ it was evident ◊ that ◊ the mathematics lecturer ◊ did not,
        ◊ as ◊ the time-consuming preparation ◊ really ◊ exhausted her.

___

        Superficial probe
        A mathematics lecturer was mentioned.

___

        Detailed probe
        A lecturer did not love an afternoon session's examples.

An additional difference in comparison to Experiment 1 is the presence of an extra clausal layer between antecedent and ellipsis. This increases the distance between the loci of encoding and integration of the antecedent, and may make subjects less likely to adopt a strategy based on memory maintenance or 'sustained reactivation' as observed by Callahan et al. (2010).

A negation occurred as part of the critical ellipsis region (*did not*) in half of the experimental items, like in (9). For the other half, the negation instead occurred in the antecedent region (*The advanced students did not love . . .*) and the critical region consisted only of the auxiliary *did*. Comprehension probes appeared after each sentence in both groups, with equal numbers of true and false statements. As in Experiment 1, subjects were required to assess the veracity of the statements given the information in the preceding sentence. Probes in the superficial group always followed the template __ *was mentioned*, featuring either an entity that appeared in the preceding sentence or an unrelated entity that had not been mentioned. Probes in the detailed group randomly targeted either the antecedent or the ellipsis, with either unchanged or reversed polarity, and sometimes with parts of the original string replaced by novel terms (*Some students loved a morning session*; correct answer: false). Other aspects of the sentences were never targeted.

3.2 Participants

Eighty-one native speakers of English recruited from the University of Massachusetts, Amherst participated in the study. Forty-one subjects were assigned to the superficial probe group, the remaining forty to the detailed probe group. All subjects received course credit for their participation, and informed consent was obtained before each experimental session. The study was approved by the local Institutional Review Board of the Linguistics Department at the University of Massachusetts, Amherst.

3.3 Procedure

The procedure was largely the same as in Experiment 1, apart from the changes to the comprehension probes described above. Instead of masked self-paced reading, as described for Experiment 1, Experiment 2 used centered self-paced reading to avoid line breaks occurring inside the antecedent region. In centered self-paced reading, each region is presented in the center of the screen and replaced with the next region when the space bar is

pressed. A fixation cross was presented for 1000 ms before each trial to mark the position of a given region's first character.

3.4 Predictions

Assuming that the overall speedup in the complex antecedent conditions of Experiment 1 was due to the use of *and*, which creates an expectation of parallelism, we should see no such effect in the current experiment, given that *but* was used instead. Should such an effect nevertheless appear, one would need to adopt a more task-oriented explanation, such as readers being anxious to get to the end of the sentence as quickly as possible. This kind of strategy might make sense if readers are afraid they might forget the information they need to answer the comprehension questions. It would then also make sense for readers in the detailed probe group to show a larger speedup, as they can expect to be queried about the sentences' contents more rigorously.

If a parallelism requirement induced by *and* was responsible for masking any antecedent complexity effects related exclusively to ellipsis processing in Experiment 1, reading times at the ellipsis site in the current study may increase, decrease or be unaffected as the antecedent becomes more complex. The first possibility would be consistent with the predictions of Murphy (1985), unless the increased distance between antecedent and ellipsis in comparison to Experiment 1 (see materials section) causes subjects to fall back on "discourse-based" processing. A decrease in ellipsis processing time for complex antecedents would support the notion of elaboration-based facilitation along the lines of Hofmeister's (2011) account. A null result, meanwhile, would lend credibility to approaches in which antecedent complexity is not expected to influence ellipsis processing at all (Frazier & Clifton 2001, 2005; Martin & McElree 2008).

On the other hand, the latter account would be called into question most strongly if the detailed probe group showed evidence of complexity effects at the point of retrieval while the superficial probe group did not. This would imply that task effects are a factor in ellipsis processing, and that the studies of Frazier & Clifton (2000) and Martin & McElree (2008) may have yielded null results due to subjects being insufficiently motivated to interpret sentences carefully.

3.5 Data analysis

Data analysis was carried out in a manner analogous to Experiment 1. The experimental factors antecedent complexity and probe type were sum-coded, with the levels "simple" and "superficial" receiving the value –1 and the levels "complex" and "detailed" receiving the value 1, respectively. Again, all models featured the maximal random effects structure, to the exclusion of a random slope for probe type by subject, since this was a between-subjects factor. As before, models were fit to individual regions of interest as well as to all the data from within two regions around the ellipsis site together.

3.6 Results
3.6.1 Question responses

Mean question response accuracy across groups and conditions was 82%. Subjects in the superficial probe group reached 90% accuracy in both conditions, while subjects in the detailed probe group reached a mean accuracy of 74% in the simple antecedent condition, as compared to 72% in the complex antecedent condition. Results show that both increased antecedent complexity ($\hat{\beta} = 0.036$, CrI: [0.018, 0.053], Pr ($\hat{\beta} > 0$) $\approx$ 1) and membership in the detailed probe group ($\hat{\beta} = 0.23$, CrI: [0.176, 0.295], Pr ($\hat{\beta} > 0$) $\approx$ 1) increased response time. There was also an interaction between the factors ($\hat{\beta} = 0.024$, CrI: [0.006, 0.044], Pr ($\hat{\beta} > 0$) $= 0.99$), such that antecedent complexity led to elevated response

times only in the detailed group (nested comparisons: $\hat{\beta} = 0.06$, CrI: [0.036, 0.085], Pr $(\hat{\beta} > 0) \approx 1$). Response accuracy was lower for the detailed group ($\hat{\beta} = -0.428$, CrI: [−0.667, −0.199], Pr $(\hat{\beta} > 0) \approx 0$) and dropped with increased response time ($\hat{\beta} = -0.644$, CrI: [−1.000, −0.306], Pr $(\hat{\beta} > 0) \approx 0$). The results are shown in Figure 7.

3.6.2 Reading times
Figure 6 plots back-transformed mean reciprocal reading times (see above) by region, along with 95% confidence intervals. Figure 8 shows the results for individual regions of
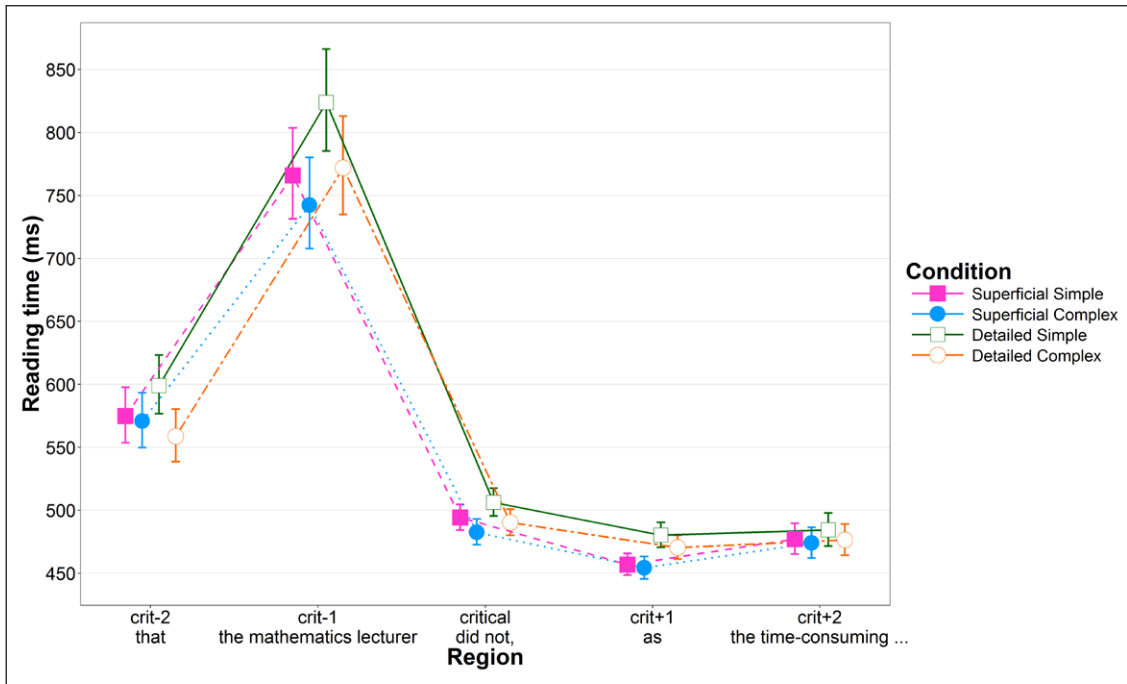
Figure 6: Experiment 2 – Mean reading times (back-transformed, on reciprocal scale) by region and condition. Error bars represent 95% confidence intervals.
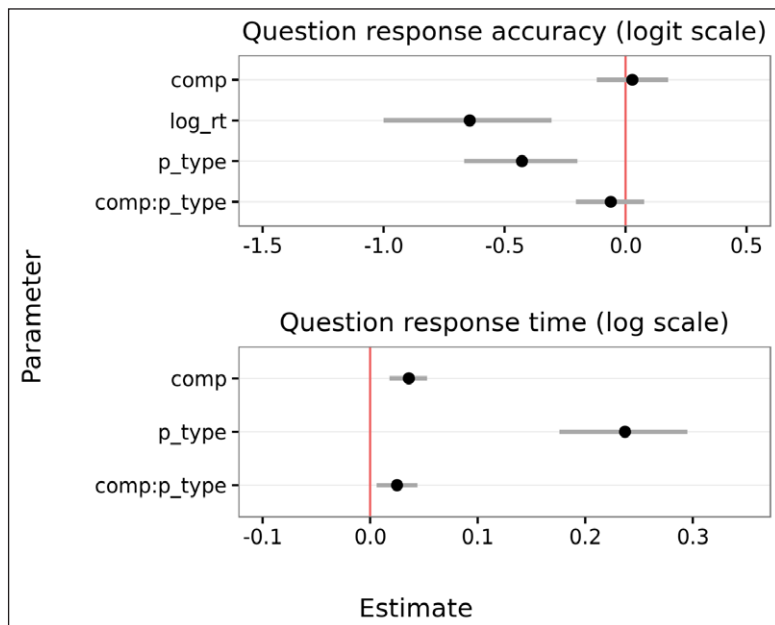
Figure 7: Experiment 2 – Question response times and accuracies: caterpillar plot of means and 95% credible intervals for parameters of interest. comp = complexity, p_type = probe type, log_rt = log response time.
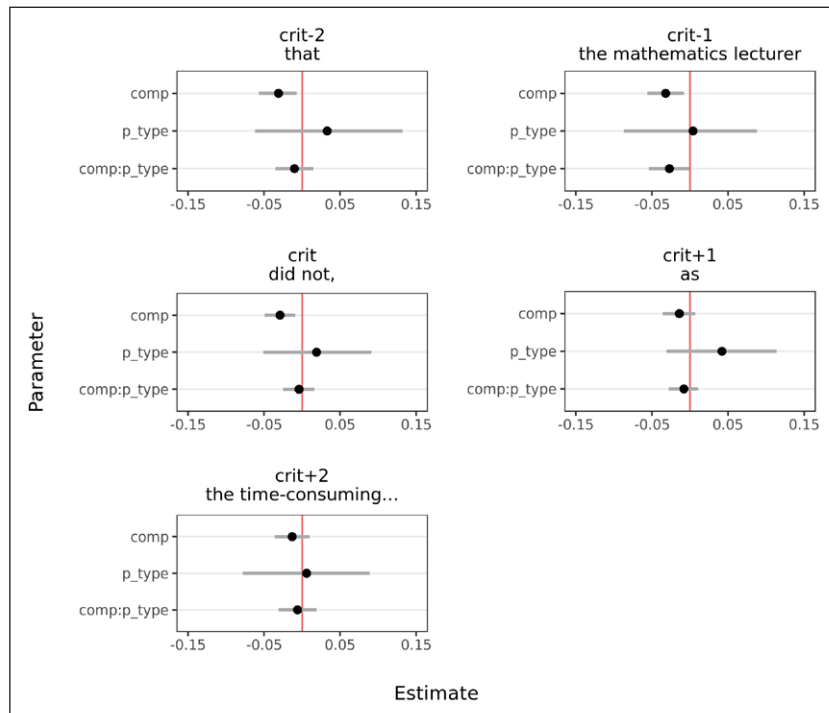
Figure 8: Experiment 2 – Caterpillar plot of means and 95% credible intervals for parameters of interest, separate models fit across regions crit–2 through crit+2. comp = complexity, p_type = probe type.

interest. At region crit–2, sentences containing complex antecedents are read more quickly ($\hat{\beta} = -0.032$, CrI: [–0.056, –0.008], Pr ($\hat{\beta} > 0$) ≈ 0), an effect which persists throughout the subsequent regions (crit–1: $\hat{\beta} = -0.031$, CrI: [–0.057, –0.007], Pr ($\hat{\beta} > 0$) = 0.01; critical region: $\hat{\beta} = -0.029$, CrI: [–0.049, –0.009], Pr ($\hat{\beta} > 0$) ≈ 0), until region crit+1, where there is only weak evidence of a speedup ($\hat{\beta} = -0.014$, CrI: [–0.036, 0.007], Pr ($\hat{\beta} > 0$) = 0.10).

The combined analysis of reading times between regions crit–2 and crit+2 shows an overall facilitatory effect of complexity ($\hat{\beta} = -0.031$, CrI: [–0.052, –0.010], Pr ($\hat{\beta} > 0$) ≈ 0), along with an interaction with probe type ($\hat{\beta} = -0.027$, CrI: [–0.049, –0.005], Pr ($\hat{\beta} > 0$) = 0.01) such that the speedup associated with more complex antecedents was larger in the detailed probe group. Probe type also influenced the changes in reading times between the critical region and region crit+1 ($\hat{\beta} = 0.026$, CrI: [–0.002, 0.054], Pr ($\hat{\beta} > 0$) = 0.96), as well as between regions crit+1 and crit+2 ($\hat{\beta} = 0.037$, CrI: [–0.068, –0.006], Pr ($\hat{\beta} > 0$) = 0.01): The negative difference between the critical region and region crit+1 was larger in the superficial group, as was the positive difference between regions crit+1 and crit+2. Model output is shown in Figure 9.

As for Experiment 1, we conducted an additional analysis based on the Bayes factor, using the same procedure as before. For the current experiment, we were particularly interested in the interaction term of the model fitted at region crit+1; this is where the probe type manipulation had an effect on reading times, but the complexity manipulation did not appear to affect processing any differently than in the other regions. The lack of a differential influence is visible in the credible intervals of the three-way interactions with the region predictor in Figure 9, which are centered roughly around zero.

Figure 10 shows the results of the Bayes factor analysis. Predictably, left-truncated prior distributions yield evidence in favor of the null hypothesis, leading to the conclusion that the sign of the interaction term is very unlikely to be positive, contra Murphy (1985). While the null hypothesis is still favored with right-truncated prior distributions, the evidence is very weak: for the two tighter priors, it is not even two times as likely as the
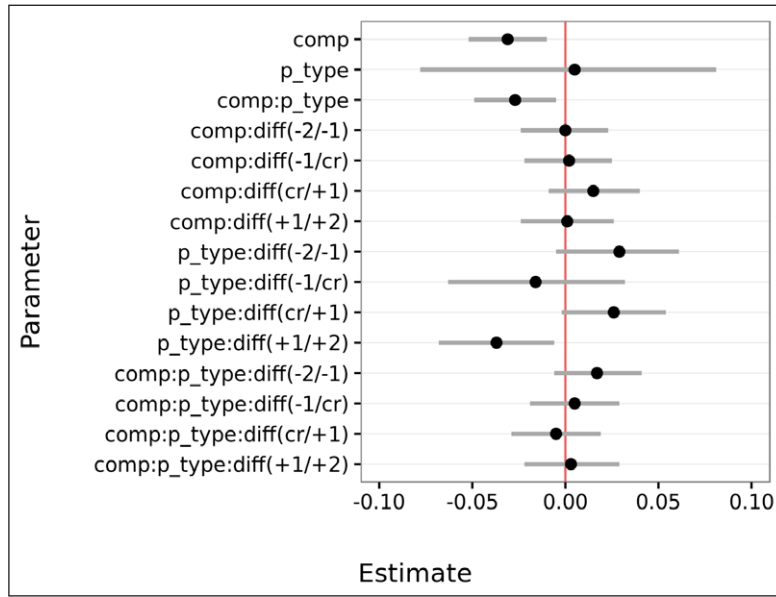
Figure 9: Experiment 2 – Caterpillar plot of means and 95% credible intervals for parameters of interest, one model fit across regions crit–2 through crit+2. *diff* parameters represent successive differences between regions, in linear order. comp = complexity, p_type = probe type.
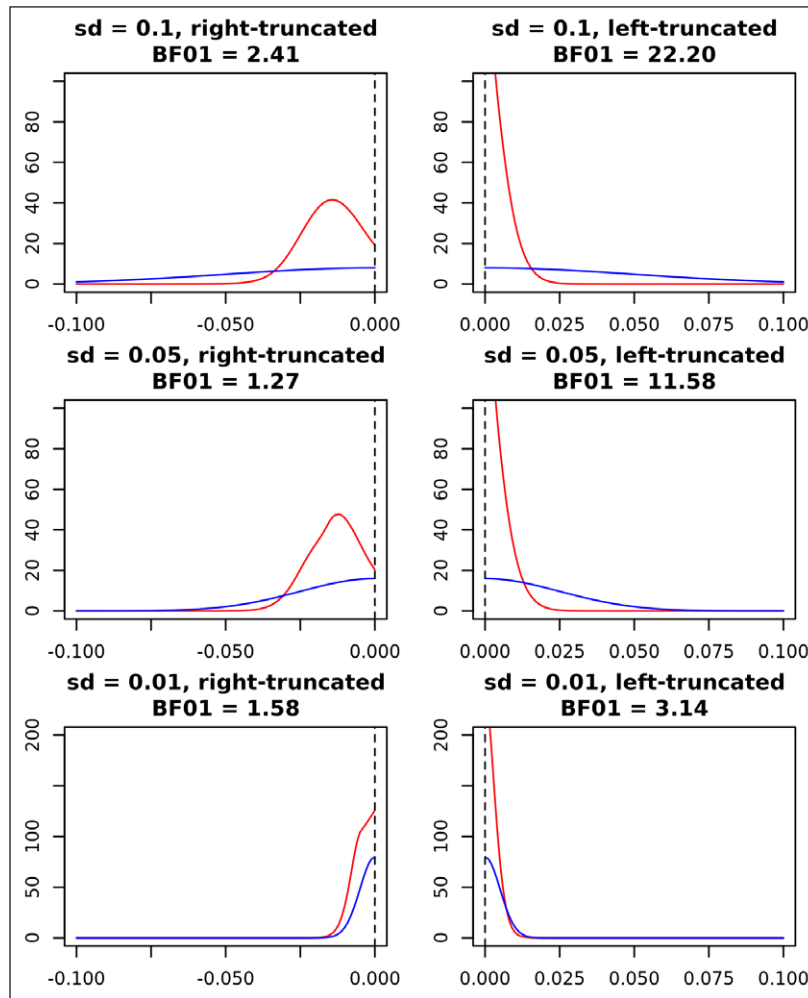
Figure 10: Experiment 2 – Density plots for the Bayes factor analysis. Blue = prior density, red = posterior sample density.

alternative. Therefore, it is possible that the greater overall speedup for sentences with complex antecedents that is visible in the detailed probe group affects region crit + 1 just like the rest of the sentence. The Bayes factor results thus yield ancillary evidence that the probe type manipulation did not interact with the antecedent complexity manipulation in a way that would support either Murphy (1985) or Hofmeister (2011), given that the interaction is either null or otherwise not limited to the predicted region.

3.7 Discussion

Given the results for the comprehension probe responses, we feel confident in claiming that our between-groups manipulation worked as intended: participants in the detailed probe group took longer to give an answer, and disproportionally longer than participants in the superficial group when the probe targeted a complex ellipsis antecedent. It thus appears that the detailed probes were indeed more difficult to answer, and that responding correctly became more difficult if information about either a more complex antecedent or a more complex ellipsis meaning was queried. However, we found no evidence of an interaction between probe type and antecedent complexity that would have been limited to critical ellipsis region. Assuming that participants in the detailed probe group processed the experimental stimuli more deeply, this result implies that the failure of earlier studies to find effects of antecedent complexity on ellipsis processing probably was not due to subjects' tendency to engage in "good enough" processing. The findings of Experiment 2 are thus in line with the predictions of pointer-based approaches, and most strongly undercut those of Murphy (1985): under Murphy's account, subjects in the detailed group would have been expected to experience a greater slowdown due to increased antecedent complexity in the critical region, given the assumption that earlier null results were due to superficial processing. We also found no evidence that would have supported the account of Hofmeister (2011), given that there was no indication of speedier retrieval of complex antecedents within as well as across groups.

As in Experiment 1, having read a more complex antecedent was associated with faster reading times for later regions. For all regions of interest taken together, the speedup interacted with the probe type manipulation, such that the reduction in overall reading times was greater for the detailed probe group. This might indicate that members of the detailed probe type were more busy trying to remember the contents of complex antecedents, and thus withdrew resources from processing. We return to this point in the general discussion.

The fact that the complexity- or length-induced speedup appeared prior to encountering the ellipsis site in Experiment 2 as well as in Experiment 1 is interesting from a methodological perspective. Remember that while the complexity manipulation introduced additional presentation regions in Experiment 1, in Experiment 2 simple and complex antecedents had the exact same number of regions. It thus seems to make no difference for the speedup effect how many presentation regions participants have passed. Rather, the quickening of the pace appears to be related to the amount of words that have been read. Keeping the number of presentation regions constant across conditions is thus not a remedy for the word-position confound that is also present in earlier studies, with the exception of Martin & McElree (2008).

The group manipulation did not appear to have any particularly strong effect on reading times for unique regions throughout the sentence, with the exception of some suggestive evidence at the region following the ellipsis. The combined analysis showed that there was a steeper drop in reading times at this position for the superficial compared to the detailed probe group, and that afterwards reading times rose more steeply for the superficial group, returning to almost identical levels across groups. It thus appears that the detailed

probe group did additional processing in this region, possibly due to spillover from the preceding ellipsis region. Indeed, the region-by-region analysis revealed suggestive evidence that the detailed probe group spent more time on region crit + 1, irrespective of antecedent complexity. Speculatively, spillover might have been a factor in Experiment 2 as opposed to Experiment 1 due to the switch to centered presentation: the latter mode may increase memory demands due to the absence of visual cues (in the form of underscores) to the surrounding linguistic context. The main effect of probe type may then be due to members of the detailed probe group allowing themselves more time to finish the antecedent-ellipsis integration, knowing that their interpretation would be queried later.

4 General discussion

We have reported two studies on antecedent complexity effects in ellipsis processing. Experiment 1 yielded evidence that increasing antecedent complexity did not influence reading times at the ellipsis site, but showed that if there is such an influence, it is unlikely to be in the form of a speedup, contra Hofmeister (2011). Similarly, the results of Experiment 2 showed no effects of antecedent complexity that would have been limited to the ellipsis site, as well as no interaction between antecedent complexity and the difficulty of the end-of-sentence probe task. Given the persistent overall speedup effect that was visible in Experiment 2, we take the results of this study to be at odds with the predictions of Murphy (1985). However, the pointer model of Frazier & Clifton (2001; 2005) and Martin & McElree (2008) is able to account for both of the findings, as the proposed memory retrieval mechanism is insensitive to antecedent complexity in terms of retrieval time.

For Experiment 1, the analysis of region-by-region data revealed that while increased antecedent complexity generally led to a decrease in reading times for the rest of the sentence, this effect was suspended at the critical region, both for ellipsis and control sentences. We have suggested that the use of *and* may have caused readers to assume parallelism between the conjuncts and created an expectation of a causal connection between the first and second clauses, leading to either maintenance or retrieval and subsequent integration of material from the first conjunct at the critical region across the board.

With regard to the account of Hofmeister (2011), our current findings show that even if certain kinds of elaboration can aid retrieval, adding genitive modifiers to object noun phrases inside VP antecedents (Experiment 2), and adverbials and modal verbs to clauses (Experiment 1), do not appear to constitute cases of such "helpful" elaboration. Whether this is a desirable corollary for the theory remains to be determined in future work.

In both studies, we observed an overall decrease in reading times in the regions between a longer, more complex antecedent and the ellipsis site. This pattern by itself is not new or surprising (Ferreira & Henderson 1993; Demberg & Keller 2008). Nevertheless, our results indicate that it does not matter in terms of the length-induced speedup if the lengthening occurs within one presentation region, as in Experiment 2, or if extra regions are added to the sentence, as in Experiment 1, which suggests that it is words, not button presses, that make people increase their reading speed over time.

If we take the difference in parallelism requirements between the conjunctions *and* and *but* seriously, the speedup also does not seem to be related to parallelism, but may have a more mundane explanation. The working memory model of Just & Carpenter (1992) assumes that sentence comprehension involves a constant trade-off between storage and processing. A reader who has already stored more information in his or her working memory will have fewer resources available to devote to the processing of incoming words. The standard view is that reading times should increase, as it takes longer to accomplish the same task with fewer resources. Given previous work on the influence of task demands on linguistic processing, however, one may ask if the reader might not

benefit from speeding up instead of slowing down. The parsing model of Lewis & Vasishth (2005), for example, assumes that linguistic information in working memory is subject to interference and decay effects which diminish the quality of the traces as new material comes in. If the participant strives to keep these traces intact, either in order to be able to answer comprehension questions or to be better able to integrate early- with late-arriving information, it may make sense to increase reading speed up to some threshold.

4.1 Conclusions

Experiment 1 yielded evidence against the assumption that increased antecedent complexity leads to faster processing of ellipsis (Hofmeister 2011). Rather, the effect of antecedent complexity is most likely null, as predicted by pointer-based accounts of ellipsis processing (Frazier & Clifton 2001, 2005; Martin & McElree 2008) or otherwise a numerically very small slowdown, as would be predicted by the account of Murphy (1985). However, the results Experiment 2 call the possibility of a slowdown into question, as no such effect became visible even when task demands were high. Still, several qualifications are in order. Ellipsis is not processed in a vacuum: sentence context and discourse relations between antecedent and ellipsis clause may enhance or mask subtle effects of complexity on retrieval, and/or interact with the manipulation themselves. It might also be that different types of antecedent complexity influence retrieval times at the ellipsis site to different degrees. Murphy's (1985) assumption of a string-copying procedure would predict the length of the antecedent to be most important, while other accounts assume that the ellipsis gap contains syntactic structure (e.g., Merchant 2001, Frazier & Clifton 2001), which would point towards factors like the number of syntactic phrases being critical. Still other accounts may claim that as ellipsis is a discourse phenomenon, and thus makes reference to a discourse model (e.g., Hardt 1993), the number of unique discourse referents contained in the antecedent may play a role. In future work, we suggest to manipulate these aspects independently in order to distinguish different theories of ellipsis processing more clearly.

Abbreviations

LKJ prior = Lewandowski-Kurowicka-Joe prior (Lewandowski et al. 2009), NP = Noun phrase, RT = Reading time, SAT = Speed-accuracy tradeoff, SE = Standard error, VP = Verb phrase

Additional File

The additional file for this article can be found as follows:

- **Appendices A and B.** Does antecedent complexity affect ellipsis processing? An empirical investigation. DOI: https://doi.org/10.5334/gjgl.290.s1

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Bader, Markus. 2014. Das Verb *lassen* im Deutschen: Zum Zusammenspiel von Auxiliarinversion und Ersatzinfinitiv [The verb *let* in German: On the interaction between auxiliary inversion and Ersatzinfinitiv]. *Zeitschrift für Sprachwissenschaft* 33(1). 1–44. DOI: https://doi.org/10.1515/zfs-2014-0001

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015b. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015a. Parsimonious mixed models. *arXiv:1506.04967*.

Box, George E.P. & David R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 26(2). 211–252.

Cai, Zhenguang G., Martin J. Pickering & Patrick Sturt. 2013. Processing verbphrase ellipsis in Mandarin Chinese: Evidence against the syntactic account. *Language and Cognitive Processes* 28(6). 810–828. DOI: https://doi.org/10.1080/01690965.2012.665932

Callahan, Sarah M., Lewis P. Shapiro & Tracy Love. 2010. Parallelism effects and verb activation: The sustained reactivation hypothesis. *Journal of Psycholinguistic Research* 39(2). 101–118. DOI: https://doi.org/10.1007/s10936-009-9128-0

Christianson, Kiel, Andrew Hollingworth, John F. Halliwell & Fernanda Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42(4). 368–407. DOI: https://doi.org/10.1006/cogp.2001.0752

Dalrymple, Mary, Stuart M. Shieber & Fernando C.N. Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy* 14(4). 399–452. DOI: https://doi.org/10.1007/BF00630923

Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2). 193–210. DOI: https://doi.org/10.1016/j.cognition.2008.07.008

Dickey, James M. & B.P. Lientz. 1970. The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics* 41(1). 214–226. DOI: https://doi.org/10.1214/aoms/1177697203

Dwivedi, Veena D. 2013. Interpreting quantifier scope ambiguity: Evidence of heuristic first, algorithmic second processing. *PLoS ONE* 8(11). e81461. DOI: https://doi.org/10.1371/journal.pone.0081461

Ferreira, Fernanda. 2003. The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47(2). 164–203. DOI: https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, Fernanda & John M. Henderson. 1993. Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale* 47(2). 247–275. DOI: https://doi.org/10.1037/h0078819

Foertsch, Julie & Morton Ann Gernsbacher. 1994. In search of complete comprehension: Getting "minimalists" to work. *Discourse Processes* 18(3). 271–296. DOI: https://doi.org/10.1080/01638539409544896

Foraker, Stephani & Brian McElree. 2011. Comprehension of linguistic dependencies: Speed-accuracy tradeoff evidence for direct-access retrieval from memory. *Language and Linguistics Compass* 5(11). 764–783. DOI: https://doi.org/10.1111/j.1749-818X.2011.00313.x

Frazier, Lyn, Alan Munn & Charles Clifton Jr. 2000. Processing coordinate structures. *Journal of Psycholinguistic Research* 29(4). 343–370. DOI: https://doi.org/10.1023/A:1005156427600

Frazier, Lyn & Charles Clifton, Jr. 2000. On bound variable interpretations: The LF-only hypothesis. *Journal of Psycholinguistic Research* 29(2). 125–140. DOI: https://doi.org/10.1023/A:1005136826534

Frazier, Lyn & Charles Clifton, Jr. 2001. Parsing coordinates and ellipsis: Copy α. *Syntax* 4(1). 1–22. DOI: https://doi.org/10.1111/1467-9612.00034

Frazier, Lyn & Charles Clifton, Jr. 2005. The syntax-discourse divide: Processing ellipsis. *Syntax* 8(2). 121–174. DOI: https://doi.org/10.1111/j.1467-9612.2005.00077.x

Frazier, Lyn, Lori Taft, Tom Roeper, Charles Clifton & Kate Ehrlich. 1984. Parallel structure: A source of facilitation in sentence comprehension. *Memory & Cognition* 12(5). 421–430. DOI: https://doi.org/10.3758/BF03198303

Gabry, Jonah & Ben Goodrich. 2016. *rstanarm: Bayesian applied regression modeling via Stan.* https://CRAN.R-project.org/package = rstanarm. R package version 2.9.0-1.

Greiner, Ben. 2015. Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association* 1(1). 114–125. DOI: https://doi.org/10.1007/s40881-015-0004-4

Grinder, John & Paul M. Postal. 1971. Missing antecedents. *Linguistic Inquiry* 2. 269–312.

Hankamer, Jorge & Ivan Sag. 1976. Deep and surface anaphora. *Linguistic inquiry* 7(3). 391–428.

Hardt, Daniel. 1993. *Verb phrase ellipsis: Form, meaning, and processing.* Philadelphia, PA: University of Pennsylvania dissertation.

Hoekstra, Rink, Richard D. Morey, Jeffrey N. Rouder & Eric-Jan Wagenmakers. 2014. Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review* 21(5). 1157–1164. DOI: https://doi.org/10.3758/s13423-013-0572-3

Hofmeister, Philip. 2011. Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes* 26(3). 376–405. DOI: https://doi.org/10.1080/01690965.2010.492642

Just, Marcel A. & Patricia A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review* 87(4). 329–354. DOI: https://doi.org/10.1037/0033-295X.87.4.329

Just, Marcel A. & Patricia A. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99(1). 122–149. DOI: https://doi.org/10.1037/0033-295X.99.1.122

Knoeferle, Pia. 2014. Conjunction meaning can modulate parallelism facilitation: Eye-tracking evidence from German clausal coordination. *Journal of Memory and Language* 75. 140–158. DOI: https://doi.org/10.1016/j.jml.2014.05.002

Kooperberg, Charles. 2016. *logspline: Logspline density estimation routines.* https://CRAN.R-project.org/package = logspline. R package version 2.1.9.

Lewandowski, Daniel, Dorota Kurowicka & Harry Joe. 2009. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 100(9). 1989–2001. DOI: https://doi.org/10.1016/j.jmva.2009.04.008

Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29(3). 375–419. DOI: https://doi.org/10.1207/s15516709cog0000_25

Martin, Andrea E. & Brian McElree. 2008. A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language* 58(3). 879–906. DOI: https://doi.org/10.1016/j.jml.2007.06.010

McElree, Brian & Johanna Nordlie. 1999. Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review* 6(3). 486–494. DOI: https://doi.org/10.3758/BF03210839

Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.

Murphy, Gregory L. 1985. Processes of understanding anaphora. *Journal of Memory and Language* 24(3). 290–303. DOI: https://doi.org/10.1016/0749-596X(85)90029-4

Nicenboim, Bruno, Pavel Logacev, Carolina Gattei & Shravan Vasishth. 2016. When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in Psychology* 7(280). DOI: https://doi.org/10.3389/fpsyg.2016.00280

Nicenboim, Bruno & Shravan Vasishth. 2016. Statistical methods for linguistic research: Foundational ideas – Part II. *Language and Linguistics Compass* 10(11). 591–613. DOI: https://doi.org/10.1111/lnc3.12207

Paape, Dario Leander Jim Felix. 2016. Filling the silence: Reactivation, not reconstruction. *Frontiers in Psychology* 7(27). DOI: https://doi.org/10.3389/fpsyg.2016.00027

Phillips, Colin & Dan Parker. 2014. The psycholinguistics of ellipsis. *Lingua* 151. 78–95. DOI: https://doi.org/10.1016/j.lingua.2013.10.003

Raftery, Adrian E. 1995. Bayesian model selection in social research. *Sociological Methodology* 25. 111–163. DOI: https://doi.org/10.2307/271063

Rayner, Keith, Gretchen Kambe & Susan A. Duffy. 2000. The effect of clause wrapup on eye movements during reading. *The Quarterly Journal of Experimental Psychology: Section A* 53(4). 1061–1080. DOI: https://doi.org/10.1080/713755934

R Core Team. 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. http://www.R-project.org/.

Rohde, Douglas. 2003. Linger, version 2.94. http://tedlab.mit.edu/~dr/Linger/.

Sorensen, Tanner, Sven Hohenstein & Shravan Vasishth. 2016. Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology* 12(3). 175–200. DOI: https://doi.org/10.20982/tqmp.12.3.p175

Stan Development Team. 2016. *Stan*. http://mc-stan.org. Version 2.9.0.

Swets, Benjamin, Timothy Desmet, Charles Clifton & Fernanda Ferreira. 2008. Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition* 36(1). 201–216. DOI: https://doi.org/10.3758/MC.36.1.201

Tanenhaus, Michael K. & Greg N. Carlson. 1990. Comprehension of deep and surface verbphrase anaphors. *Language and Cognitive Processes* 5(4). 257–280. DOI: https://doi.org/10.1080/01690969008407064

Wagenmakers, Eric-Jan, Tom Lodewyckx, Himanshu Kuriyal & Raoul Grasman. 2010. Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology* 60(3). 158–189. DOI: https://doi.org/10.1016/j.cogpsych.2009.12.001

Warren, Tessa, Sarah J. White & Erik D. Reichle. 2009. Investigating the causes of wrap-up effects: Evidence from eye movements and E–Z Reader. *Cognition* 111(1). 132–137. DOI: https://doi.org/10.1016/j.cognition.2008.12.011

Williams, Edwin S. 1977. Discourse and logical form. *Linguistic Inquiry* 8(1). 101–139.

Xiang, Ming, Julian Grove & Jason Merchant. 2014. Ellipsis sites induce structural priming effects. Ms, University of Chicago.