RESEARCH

A memory-based explanation of antecedent-ellipsis mismatches: New insights from computational modeling

Dan Parker

College of William and Mary, Williamsburg, VA 23185, US
dparker@wm.edu

An active question in psycholinguistics is whether or not the parser and grammar reflect distinct cognitive systems. Recent evidence for a distinct-systems view comes from cases of ungrammatical but acceptable antecedent-ellipsis mismatches (e.g., *Tom kicked Bill, and Matt was ~~kicked by Tom~~ too.). The finding that these mismatches show varying degrees of acceptability has been presented as evidence for the use of extra-grammatical parsing strategies that restructure a mismatched antecedent to satisfy the syntactic constraints on ellipsis (Arregui et al. 2006; Kim et al. 2011). In this paper, I argue that it is unnecessary to posit a special class of parser-specific rules to capture the observed profiles, and that acceptable mismatches do not reflect a parser-grammar misalignment. Rather, such effects are a natural consequence of a single structure-building system (i.e., the grammar) that relies on noisy, domain-general memory access mechanisms to retrieve an antecedent from memory. In Experiment 1, I confirm the acceptability profiles reported in previous work. Then in Experiment 2, as proof-of-concept, I show using an established computational model of memory retrieval that the observed acceptability profiles follow from independently motivated principles of working memory, without invoking multiple representational systems. These results contribute to a uniform memory-based account of acceptable ungrammaticalities for a wide range of dependencies.

1 Introduction

Ellipsis has received much attention in formal linguistics, but relatively little is known about how speakers build and interpret antecedent-ellipsis dependencies during real-time comprehension. The current study bridges this gap by investigating a class of apparent mismatches between the ellipsis structures constructed during real-time comprehension and those licensed by the grammar.

Many formal accounts assume that the antecedent and ellipsis must match in syntactic form, typically stated in terms of a formal syntactic identity constraint (e.g., Sag 1976; Williams 1977; Lappin 1992; 1996; Fiengo & May 1994). However, the acceptability of antecedent-ellipsis mismatches in examples like (1a–d) challenges this view (Dalrymple et al. 1991; Hardt 1993; Lascarides & Asher 1993; Kehler 2000). For instance, (1a) involves an active ellipsis clause with a passive antecedent, yet there appears to be little to no degradation in acceptability resulting from this syntactic mismatch.

(1)   a.   This information could have been released by Gorbachov, but he chose not to ~~release this information~~.
       b.   This problem was to have been looked into, but obviously nobody did ~~look into this problem~~.
       c.   Of course this theory could be expressed using SDRSs, but for the sake of simplicity we have chosen not to ~~express this theory using SDRSs~~.
       d.   In March, four fireworks manufacturers asked that the decision be reversed, and on Monday the ICC did ~~reverse the decision~~.

The acceptability of antecedent-ellipsis mismatches like those in (1) raises important questions about how mismatches are represented mentally and how they are processed in real time, bringing into light the concern over whether experimental and computational evidence can be brought to bear on these questions (see Phillips & Parker 2014, for discussion). To date, only a handful of studies have attempted to address these questions using experimental methods, which are based mostly on acceptability rating measures to determine the grammatical status of antecedent-ellipsis mismatches (Arregui et al. 2006; Kim et al. 2011; Kertz 2013; Kim & Runner 2018). Although the debate over the grammatical status is not yet resolved, these studies present converging evidence for a cline of acceptability across different types of antecedent-ellipsis mismatches (Arregui et al. 2006; Kim et al. 2011; Kim & Runner 2018). Specifically, these studies have shown that even when a syntactically matching antecedent is not available, speakers can still resolve ellipsis to varying degrees of acceptability, such that acceptability declines as the degree of the syntactic mismatch between the antecedent and ellipsis increases.

For instance, in an offline acceptability judgment study, Arregui and colleagues (2006) found that a passive-active mismatch, e.g., (2a), was judged as more acceptable than an active-passive mismatch, e.g., (2b). They also found that VP ellipsis with an antecedent in a verbal gerund, e.g., (3a), was judged as more acceptable than maximally similar sentences with an antecedent in a nominal gerund, e.g., (3b). Kim and colleagues (2011; 2018) provide converging evidence for these differences with data from a series of magnitude estimation experiments. They also extended the acceptability cline by showing that passive-passive matches are less acceptable than active-active matches. Broadly, these studies show the following pattern: matching voice > gerundive antecedents > mismatching voice (> indicates "more acceptable than").

(2)   **Mismatching voice:** Passive-Active > Active-Passive
       a.   The dessert was praised by the customer after the critic did already.
       b.   The customer praised the dessert after the appetizer was already.

(3)   **Gerundive antecedents:** Verbal Gerund > Nominal Gerund
       a.   Singing the arias tomorrow night will be difficult, but Maria will.
       b.   Tomorrow night's singing of the arias will be difficult, but Maria will.

Arregui et al. (2006) and Kim et al. (2011) make opposite claims about the grammatical status of antecedent-ellipsis mismatches but agree that the acceptability cline can be explained by extra-grammatical processing constraints or additional processing heuristics. Arregui et al. (2006) argue that antecedent-ellipsis mismatches are ungrammatical, under the assumption that syntactic identity is based on a surface-true representation of the sentence, and that when the antecedent does not match the elided constituent, the parser repairs ("recycles") the antecedent to create a syntactically matching form using the same processing operations that are applied for garden path recovery. According to this account, the acceptability cline reflects the amount of repair work required to restruc-

ture the antecedent into a syntactically matching form, such that more repair work leads to lower acceptability. For example, the passive-active mismatch in (2a) is predicted to be more acceptable than the active-passive mismatch in (2b) because it requires less work to recover an active form from a more complex passive antecedent than restructuring an active antecedent into passive form (drawing on insights from early recall studies, cf. Mehler 1963). Likewise, for the gerundive antecedents, (3a) is predicted to more acceptable than (3b), because it is easier to restructure a verbal gerund into a matching antecedent than it is to create one from scratch with a nominal gerund.

Kim et al. (2011) also adopt a syntactic analysis of mismatches, but argue that mismatches are grammatical, assuming a more flexible relation between surface forms. They propose an account in which a canonical VP is part of the syntactic representation of all the antecedents in (2) and (3), but the correct, matching form is obscured by further steps in the syntactic derivation of the surface form. On this view, the acceptability cline reflects the number of derivational steps or parser states that the processor must search through to recover a syntactically matching antecedent, such that more search work leads to lower acceptability. Importantly, the search process does not proceed at random, but rather is guided by a set of ellipsis-specific parsing heuristics that prioritize certain structures in the search space, such as structures with in-situ arguments (e.g., no passivization) and maximal ellipsis (e.g., the MaxElide constraint described in Takahashi & Fox 2005). According to this account, ellipsis with mismatching voice, e.g., (2), is predicted to be less acceptable than ellipsis with matching voice, because a voice mismatch violates the preference for structures with in-situ arguments, and hence requires more search work to recover a matching antecedent.

More recently, Kertz (2013) argued that the gradient acceptability observed for voice mismatches reflects a violation of information-structural constraints, rather than the violation of extra-grammatical processing constraints. Kertz noted that the accounts proposed by Arregui et al. (2006) and Kim et al. (2011) incorrectly predict comparable acceptability for the sentences in (4), since the antecedents and elided verb phrases are identical in both (4a) and (4b). According to Kertz, the degraded acceptability of (4a) relative to (4b) reflects a violation of an information-structural constraint governing contrastive topics. Specifically, when the subject/topic of the ellipsis clause, e.g., "the pedestrian", is in focus, it must be interpreted contrastively with an argument in the antecedent clause, e.g., "the driver", as it is in (4b). But this constraint is violated in (4a), since "the driver" appears in a non-topical position. This account suggests that additional factors such as information structure may further impact acceptability. However, it is unclear how this account might extend to other types of mismatches, such as those involving gerundive antecedents, e.g., (3), where the antecedent appears in a matching subject/topic position. Furthermore, it is unclear whether the proposed constraints on information structure are available to guide initial ellipsis resolution, or whether they are even applied during real-time ellipsis resolution (see Kertz 2013 for discussion).

(4)   a. #The incident was reported by the driver, and the pedestrian did ~~report the incident~~ too.
        b. The incident was reported by the driver, although he didn't really need to ~~report the incident~~.

Previous accounts of the acceptability cline inform theories of language processing in three ways. First, existing findings are relevant for questions concerning the nature of the linguistic representation, such as whether the antecedent-ellipsis relationship requires syntactic or semantic identity. Acceptable antecedent-ellipsis mismatches have previously

been presented as evidence for a semantic identity constraint (Dalrymple et al. 1991; Hardt 1992; 1993; 1999a; b; Kehler 1993a; Ginzburg & Sag 2000; Merchant 2001; van Craenenbroeck 2010), but speakers' sensitivity to subtle syntactic distinctions has been taken to suggest that ellipsis requires a syntactically matching antecedent (see Kim et al. 2011 and Kim & Runner 2018 for discussion). Second, existing accounts inform our understanding of real-time language processes. In particular, the acceptability cline has been presented as evidence for a special class of extra-grammatical, parser-specific heuristics that reanalyze, repair, or restructure ungrammatical antecedents to meet the syntactic constraints on ellipsis. Third, existing accounts inform our understanding of the processing architecture. If ellipsis resolution relies on parsing heuristics that are not part of the grammar proper, then this amounts to the claim that the grammar and parser reflect distinct structure-building systems, reinforcing the grammar-parser distinction that has been the standard assumption in linguistics and psycholinguistics since the 1970s (Fodor et al. 1974; Levelt 1974; Townsend & Bever 2001; Ferreira, Bailey & Ferraro 2002; Ferreira & Patson 2007).[1]

1.1 Revisiting previous conclusions about antecedent-ellipsis mismatches

There are two reasons to revisit previous conclusions about the ellipsis acceptability cline. First, it is difficult to distinguish the accounts proposed by Arregui et al. (2006) and Kim et al. (2011). Both accounts appeal to parser operations to capture the gradient acceptability, and without any strongly divergent predictions, they cover much of the same empirical ground (see Kim & Runner 2018 for discussion).

Second, the experimental findings for antecedent-ellipsis mismatches resemble findings from previous studies on other syntactic dependencies involving anaphora and agreement, where different conclusions are drawn about the nature of the underlying representations, mental processes, and processing architecture. For instance, many studies have shown that ungrammatical subject-verb agreement relations in configurations like (5) are often overlooked during real-time language processing, due to the presence of a non-subject plural lure, e.g., "the cabinets", leading to increased acceptability relative to sentences that lack a plural item (e.g., Pearlmutter et al. 1999; Wagers et al. 2009). This effect is described in the literature as "agreement attraction", and constitutes a kind of "acceptable ungrammaticality", in the same sense that antecedent-ellipsis mismatches constitute a case of acceptable ungrammaticality (Arregui et al. 2006).

(5)   *The key to the cabinets unsurprisingly were rusty after years of disuse.

Many researchers have argued that the mistaken acceptability of sentences like (4) arises not because of the application of parser-specific heuristics, extra-grammatical processes, or separate structure-building systems for the grammar and parser, but rather because the grammatical constraints on syntactic dependency formation are implemented in real time using a noisy memory retrieval system that creates the opportunity for errors and gradient acceptability (Wagers et al. 2009; Dillon et al. 2013; Tanner et al. 2014; Lago et al. 2015;

---

[1] Arregui et al. argue that the processor repairs mismatches using the same mechanisms that are used for garden-path recovery, according to operations defined by the grammar. As such, the proposed repair rules would not constitute motivation for a parser-grammar distinction. However, the repair rules for antecedent-ellipsis mismatches and garden path sentences must be qualitatively different. For garden path sentences, the parser pursues an alternative structural analysis that is licensed by the input. For ellipsis, the parser radically changes the input to create a matching antecedent, sometimes creating one from scratch, e.g., (3b). That is, the parser creates a representation that is not licensed by the input. There are cases where noisy or ambiguous input licenses repair, but there is little reason to believe that the listener failed to properly encode the input in sentences like (2)–(3). As such, it is unclear how such radical changes to the encoding would be licensed by the grammar in the case of antecedent-ellipsis mismatches.

Tucker et al. 2015; Parker & Phillips 2016; 2017). On this view, long-distance syntactic dependencies such as subject-verb agreement are implemented in real time by retrieving a licensor from memory using a cue-based retrieval mechanism (McElree 2000; 2006; McElree et al. 2003; Lewis & Vasishth 2005; Lewis et al. 2006). This mechanism works by probing previously processed material for a constituent that matches a particular set of feature-based retrieval cues, such as $+s$ubject and $+p$lural for subject-verb agreement, to satisfy the structural and featural requirements on the dependency. In sentences like (5), a partial match to the non-subject noun "cabinets" *b*ased on the $+p$lural *f*eature can give the impression that agreement is licensed, boosting acceptability of an otherwise ungrammatical sentence (Wagers et al. 2009).

This retrieval-based account of acceptable but ungrammatical subject-verb agreement relations has been extended to a wide range of acceptable ungrammaticalities involving case licensing (Sloggett 2013), reflexive licensing (Patil et al. 2016; Parker & Phillips 2017), negative polarity item licensing (Vasishth et al. 2008), and even agreement in noun phrase (NP) ellipsis (Martin et al. 2012; 2014), but not yet cases of acceptable ungrammaticalities involving VP-ellipsis. Crucially, in each of these studies, the acceptable ungrammaticalities are not taken as evidence for a special class of parser-specific rules or distinct structure-building systems for the grammar and parser. Rather, they all assume that parsing relies on a single structure-building system (i.e., the grammar) implemented in a noisy general cognitive architecture. Based on these findings, taken together with the growing evidence that ellipsis is resolved in real-time using the same error-prone retrieval mechanism proposed for these other dependencies (Martin & McElree 2008; 2009; 2011; Martin, et al. 2012; 2014), the question arises whether acceptable ungrammaticalities involving ellipsis can be captured in a similar fashion.

1.2 The present study

The present study tests whether the VP ellipsis acceptability cline can be captured under the same retrieval-based account that has been proposed for the wide range of acceptable ungrammaticalities observed for other family-related dependencies. On this view, the acceptability cline observed for antecedent-ellipsis mismatches does not reflect the application of parser-specific heuristics or a parser-grammar misalignment. Rather, such effects are a natural consequence of a single structure-building system that relies on noisy memory retrieval mechanisms to recover an antecedent. According to this account, when a comprehender encounters an ellipsis site, they engage a retrieval process to find an antecedent that matches the structural and morphological retrieval cues (i.e., the search criteria) compiled at the ellipsis site. When an antecedent does not match the search criteria in some way, as in the case of mismatching antecedent-ellipsis sentences, a processing disruption is observed, which can be detected in acceptability ratings. Based on previous findings on the relation between memory retrieval and acceptability (e.g., Dillon et al. 2015), the size of that disruption is determined by the degree of match between the retrieval cues and the corresponding features of the antecedent (i.e., "probe-to-target" similarity), such that antecedents with a greater mismatch are more difficult to interpret and consequently rated as less acceptable than their matching counterparts.

This proposal is developed in two experiments in the current study. To preview, Experiment 1 uses untimed acceptability judgments to confirm the acceptability cline reported in previous studies (Arregui et al. 2006; Kim et al. 2011). Then, as proof-of-concept, Experiment 2 uses an established computational model of memory retrieval (ACT-R; Lewis & Vasishth 2005) to show that the observed acceptability profiles follow from independently motivated principles of working memory, without invoking multiple structure-building systems. The model accurately predicts the behavioral profiles observed

in Experiment 1 as a consequence of probe-to-target similarity, without adjusting the model's default parameters.

2 Experiment 1: Confirmation of the ellipsis acceptability cline

The VP ellipsis acceptability cline has clear implications for our understanding of the relationship between linguistic constraints and real-time processing mechanisms. However, it is difficult to compare acceptability profiles between previous studies (e.g., Arregui et al. 2006; Kim et al. 2011; Kertz 2013), as they used different experimental designs, methodologies, and materials. To address this issue, Experiment 1 provides a within-participants comparison of the mismatches common between previous studies using a single acceptability measure.

2.1 Participants

Participants were 36 native speakers of English who were recruited using Amazon's Mechanical Turk web service (https://www.mturk.com). This sample size was determined by a statistical power test generated using previously published results. Given the baseline .73-point difference between voice mismatches reported in Arregui et al. (2006), the power analysis suggested that 35 participants would be needed to achieve power of 80%. One additional participant for a total of 36 was recruited to balance list presentation. Participation required an IP address in the United States, and each participant was screened for native speaker abilities. The screening probed participants' knowledge of the constraints on English tense, modality, morphology, and syntactic islands. Participants were required to meet the following qualifications: location was restricted to the United States, HIT approval rate for all requesters' HITs was greater than or equal to 98, and the number of HITs approved for each participant was greater than or equal to 5000. Participants were assigned a qualification after they completed the study that would prevent them from participating in the experiment more than once. The experiment lasted approximately 20 minutes, and each participant received $3 for participating in the experiment.

2.2 Materials

Materials consisted of 12 item sets of VP ellipsis sentences with matching voice (active-active and passive-passive), 12 item sets of sentences with mismatching voice (active-passive and passive-active), and 12 item sets of sentences with gerundive antecedents (verbal and nominal gerunds), as shown in Table 1. These item sets were created in a way to intersect all previous experimental studies on ellipsis mismatches (Arregui et al. 2006; Kim et al. 2011; Kertz 2013 Kim & Runner 2018). All materials were taken directly from the Appendices of Arregui et al. (2006). The full materials list is provided in the Supple-

Table 1: Example items for Experiment 1.

Voice match
Active-Active: Jill betrayed Abby, and Matt did too.
Passive-Passive: Abby was betrayed by Jill, and Matt was too.
Voice mismatch
Passive-Active: Jill was betrayed by Abby, and Matt did too.
Active-Passive: Abby betrayed Jill, and Matt was too.
Gerundive antecedent
Verbal Gerund: Singing the arias tomorrow night will be difficult, but Maria will.
Nominal Gerund: Tomorrow night's singing of the arias will be difficult, but Maria will.

mentary Materials. The sentences from each item set were distributed in a Latin square design across 2 lists, such that each participant read 6 sentences for each (mis)match type.

2.3 Procedure

Sentences were presented using Ibex (http://spellout.net/ibexfarm/). Participants were instructed to rate the acceptability of each sentence using a 7-point scale (7 = most acceptable, 1 = least acceptable). Participants were required to complete the experiment in one hour, which gave them adequate time to rate each sentence. All participants completed the task within the time limit. Each sentence was displayed in its entirety on the screen along with the rating scale. Participants could click boxes or use the numerical keypad to enter their ratings. The order of presentation was randomized for each participant.

2.4 Data analysis

Data were analyzed using linear mixed-effects models, with fixed factors for experimental manipulations and a fully specified random effects structure, which included random intercepts and slopes for all fixed effects by participants and by items (Baayen et al. 2008; Barr et al. 2013). Each model included simple effect-coded fixed effects ($+0.5/-0.5$ for each level within each mismatch type). To ensure that all comparisons were made, three additional models were coded to test for differences across mismatch types (voice match vs. voice mismatch, voice match vs. gerundive antecedents, voice mismatch vs. gerundive antecedents). Models were built using the lme4 package (Bates et al. 2018) in the R software environment (R Development Core Team 2014). If there was a convergence failure or if the model converged but the correlation estimates were high, the random effects structure was simplified following Baayen et al. (2008) until convergence obtained. An effect was considered significant if the absolute t-value was greater than 2 (Gelman & Hill 2007).

2.5 Results

The mean acceptability ratings by condition for Experiment 1 are presented in Figure 1. Figure 2 provides the histograms of ratings for each condition. Overall, matching voice sentences were rated higher than mismatching voice sentences and gerundive antecedent sentences (matching voice vs. mismatching voice: $\hat{\beta} = 1.49$, $SE = 0.12$, $t = 12.24$; matching voice vs. gerunds: $\hat{\beta} = 0.91$, $SE = 0.23$, $t = 3.90$), and sentences with a gerundive anteced-
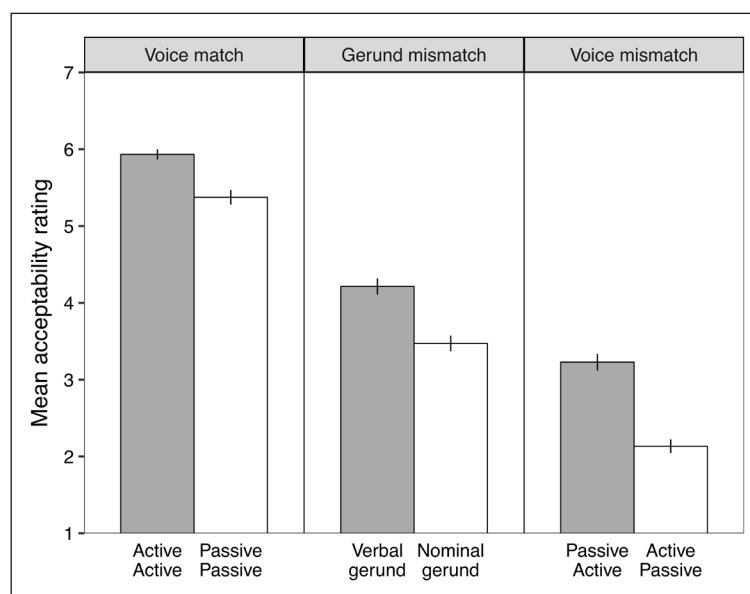

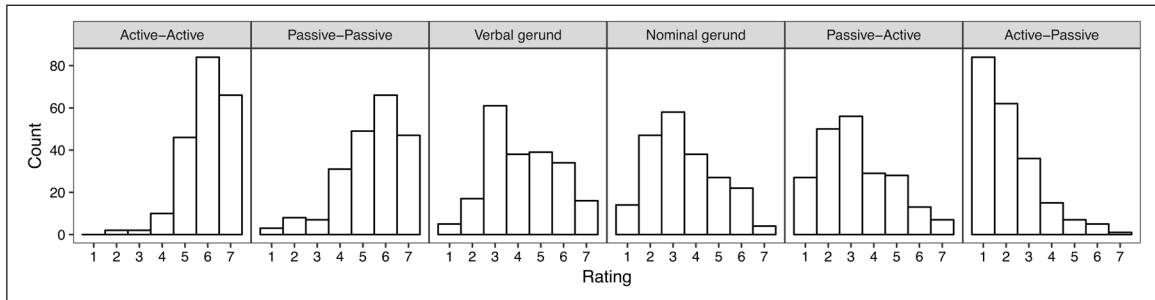**Figure 1:** Mean acceptability ratings and standard errors by participants for Experiment 1.

Figure 2: Histograms of ratings for Experiment 1.

ent were rated higher than mismatching voice sentences ($\hat{\beta} = 0.58$, $SE = 0.25$, $t = 2.27$). Within the voice match sentences, active-active forms were rated higher than passive-passive forms ($\hat{\beta} = 0.56$, $SE = 0.13$, $t = 4.30$). Within the voice mismatch sentences, passive-active forms were rated higher than active-passive forms ($\hat{\beta} = 1.11$, $SE = 0.18$, $t = 6.11$). There was also a significant interaction between voice match and voice ($\hat{\beta} = -0.55$, $SE = 0.17$, $t = -3.17$). Lastly, within the gerundive antecedent sentences, verbal gerunds were rated higher than nominal gerunds ($\hat{\beta} = 0.82$, $SE = 0.20$, $t = 3.92$).

2.6 Discussion

The results of Experiment 1 replicated each portion of the acceptability cline reported in previous studies using a single, uniform acceptability measure across (mis)match types. Previous studies based on Likert scale judgments (Arregui et al. 2006) and magnitude estimation judgments (Kim et al. 2011) showed that voice matches are more acceptable than voice mismatches and sentences with gerundive antecedents, and that sentences with gerundive antecedents are more acceptable than voice mismatches (i.e., matching voice > gerundive antecedents > mismatching voice). Furthermore, active-active forms were shown to be more acceptable than passive-passive forms, sentences with a verbal gerund antecedent were shown to be more acceptable than sentences with a nominal gerund antecedent, and passive-active forms were shown to be more acceptable than active-passive form. Experiment 1 confirmed each of these contrasts. The results of Experiment 1 provide a single, uniform measure of the acceptability cline that will serve as the basis for the computational model of ellipsis resolution described in the next section.

A concern with the results of Experiment 1 is whether the observed effects are specific to elliptical structures. For instance, the observed contrasts may simply reflect an identity contrast for coordinated structures, or a general requirement on information-structural coherence (e.g., Kertz 2013), with VP-ellipsis showing one instance of these constraints in action. However, recent work by Kim and colleagues (Kim et al. 2011; Kim & Runner 2018) has shown that the contrasts are specific to ellipsis. Kim and colleagues reasoned that if the contrasts observed for voice and nominal mismatches reflect general well-formedness constraints that are not specific to ellipsis, then the same effects should be equally present for the non-elliptical counterparts. Across several experiments, Kim and colleagues consistently found that the degradation in acceptability for voice and nominal mismatches was limited to VP-ellipsis structures. These findings suggest that the observed contrasts reflect an ellipsis-specific structural identity condition, which sets the stage for the memory-based account proposed in the next section.

3 Computational model of antecedent-ellipsis mismatches

This section describes how the acceptability cline observed in Experiment 1 can be captured under a memory-based account in a single structure-building architecture. A growing number of studies suggest that ellipsis is resolved in real-time by retrieving an antecedent

using a cue-based retrieval mechanism (Martin & McElree 2008; 2009; 2011; Martin et al. 2012; Paape 2016). Importantly, this is the exact same mechanism that has been argued to underlie many other cases of "acceptable ungrammaticalities", such as those observed for agreement, anaphora, case licensing, and negative polarity item licensing (Vasishth et al. 2008; Wagers et al. 2009; Dillon et al. 2013; Sloggett 2013; Tanner et al. 2014; Lago et al. 2015; Tucker et al. 2015; Parker & Phillips 2016; 2017; Patil et al. 2016). I argue that the acceptability cline observed for VP ellipsis is also a product of this noisy cue-based retrieval mechanism.

On this account, encountering an ellipsis site triggers a retrieval process that seeks a matching antecedent using a set of structural and morphological cues compiled at the ellipsis site. When the antecedent does not match the search criteria, such as in the case of an antecedent-ellipsis mismatch, a processing disruption is observed, resulting in decreased acceptability (e.g., Experiment 1). Based on previous findings on the relation between memory retrieval and acceptability (e.g., Dillon et al. 2015), the size of the disruption is monotonically related to the degree of match between the retrieval cues and the corresponding features of the antecedent (a relationship referred to as "probe-to-target" similarity or "cue diagnosticity" in the memory literature; Nairne 2002a), such that acceptability decreases as the degree of mismatch increases. According to this account, the gradient acceptability of VP ellipsis mismatches does not reflect the violation of parsing heuristics or the application of special repair rules, as previously claimed (e.g., Arregui et al. 2006; Kim et al. 2011), but rather violation of the expectation for what the antecedent should look like, according to the retrieval cues specified by the grammar at the ellipsis site.

To test this proposal, Experiment 2 simulated the processing of the (mis)matching ellipsis forms in Table 1 using an established model of memory retrieval, specifically, Lewis and Vasishth's (2005) ACT-R cue-based model of sentence processing [adapting code originally written by Badecker and Lewis 2007]. ACT-R (Adaptive Control of Thought—Rational; Anderson et al. 2004) is a general cognitive architecture based on independently motivated principles of memory and cognition and has been used to study a wide range of cognitive behavior involving memory access, attention, executive control, reasoning, decision making, and learning. The ACT-R model of sentence processing applies the core cognitive principles embodied in the ACT-R framework to the task of sentence processing. Importantly, the ACT-R model of sentence processing has been shown to provide a good fit to a wide range of behavioral data, including acceptability judgments (e.g., Lewis & Vasishth 2005; Vasishth et al. 2008), and thus provides a suitable model to simulate the relation between retrieval and acceptability for ellipsis resolution.

3.1 Model details

A computationally complete model of retrieval for ellipsis resolution must specify (i) the fixed mechanisms and principles of the memory architecture, (ii) the cues that drive retrieval, (iii) the antecedent structures encoded in memory that those cues target, and (iv) the linking hypothesis that relates the principles and mechanisms of retrieval to behavioral measures. These components are described in detail in the next four subsections.

3.1.1 Memory access mechanisms

In the ACT-R model of sentence processing, the words and phrases of a sentence are encoded as "chunks" (Miller 1956) in content-addressable memory (Kohonen 1980), and the hierarchical structure of the sentence is represented using pointers that index the relations between chunks. Chunks are encoded as bundles of feature-value pairs, which are inspired by the attribute-value matrices described in head-driven phrase structure grammars (e.g., Pollard & Sag 1994). Features are specified for lexical content (e.g., morpho-

syntactic and semantic features), syntactic information (e.g., category, case), and local hierarchical relations (e.g., parent, daughter, sister). Values for features include symbols (e.g., ± singular, ± animate) or pointers to other chunks (e.g., NP_1, VP_2).

Long-distance dependencies, such as those involving ellipsis, subject-verb agreement or anaphora, are formed using a domain-general cue-guided retrieval mechanism that probes all previously encoded chunks in memory in parallel to recover the head of the dependency (i.e., the target/antecedent/licensor) using a set of retrieval cues that are compiled into a single retrieval probe at the retrieval site. Retrieval cues are derived from the current word, the linguistic context, and grammatical constraints, and correspond to a subset of the structural, morphological, and semantic features of the antecedent (Lewis et al. 2006). When retrieval is engaged, potential antecedents are differentially activated based on their match to the retrieval cues. In ACT-R, antecedents become more highly activated as they match more cues, leading to a faster retrieval latency (i.e., the amount of time it takes to retrieve the antecedent from memory for further processing) and boosted acceptability (i.e., antecedents that are more highly activated are easier to process and hence more acceptable). Conversely, antecedents that mismatch the retrieval cues will have a relatively lower activation, leading to slower retrieval latencies and degraded acceptability.

These features distinguish ACT-R from other models of cue-based retrieval in two ways. First, not all theories of cue-based retrieval assume that a search process (whether parallel, as assumed in ACT-R, or serial, as assumed in traditional models of working memory, e.g., Sternberg 1975) is required to access an antecedent for non-adjacent syntactic dependencies. For instance, work by McElree and colleagues (McElree & Dosher 1989; 1993; McElree 2000; 2006; McElree et al. 2003) using speed-accuracy tradeoff (SAT) measures, and earlier work by Murdock (1971), has shown that the cues used in retrieval provide direct access to the relevant working memory representations without the need to search through irrelevant memory representations to find the antecedent. Crucially, the results of the current study do not depend on whether items in memory are accessed in parallel or directly, as the ellipsis structures tested in the current study involve only a single antecedent that is targeted at retrieval.

Second, ACT-R assumes that items with a higher activation have a higher probability of retrieval and are retrieved from memory more quickly (described in ACT-R as "retrieval latency"), but the assumption that retrieval speed is modulated by an item's activation is specific to the ACT-R implementation. For instance, McElree and colleagues have consistently shown that retrieval speed is constant or time invariant, regardless of the number of items in memory or the linear and hierarchical distance between the antecedent and retrieval site. On this view, differences in activation modulate the amount of time it takes to integrate an item back into the current sentential context (i.e., by modifying the retrieved item with a feature reflecting its downstream dependency), but they do not affect retrieval speed (see Parker et al. 2017 for discussion).

To reconcile these differences, I adopt the more general proposal that all accounts agree on, namely that higher activation at retrieval leads to faster processing (measured in terms of either retrieval or integration speed) and boosted acceptability. Throughout, I will use "processing time" to refer to timing differences in an implementation neutral way, but adopt ACT-R's principle of retrieval latencies for modeling purposes because it provides a mathematically precise expression of processing speed. In the current study, the inverse relation between antecedent activation and processing time plays a central role in explaining the acceptability cline for antecedent-ellipsis mismatches.

In ACT-R, the activation of an antecedent A_i is defined according to Equation 1, which makes explicit four principles that are known to impact memory access: (i) an item's

baseline activation B_i at the time of retrieval, (ii) the match between the item and each of the j retrieval cues in the retrieval probe S_{ji}, (iii) the penalty for a partial match PM between the cues of the retrieval probe and the item's feature values (i.e., a penalty for matching some, but not all of the cues), and (iv) stochastic noise.[2]

(6)   Equation 1

$$A_i = B_i + \sum_{j=1}^{m} W_j S_{ji} - \sum_{k=1}^{p} PM_{ki} + \epsilon$$

Baseline activation B_i is calculated according to Equation 2, which describes the usage history of chunk i as the summation of n successful retrievals of i, where t_j reflects the time since the jth successful retrieval of i to the power of the negated decay parameter d. The output is passed through a logarithmic transformation to approximate the log odds that the chunk will be needed at the time of retrieval, based on its usage history. After a chunk has been retrieved, the chunk receives an activation boost, followed by decay. However, it should be noted that the notion of decay as an explanatory concept in the memory literature is debated (e.g., Peterson & Peterson 1959; Nairne 2002b; McElree 2006; Berman et al. 2009; Lewandowsky et al. 2009). Crucially, the current results do not hinge on the decay function of Equation 2.

(7)   Equation 2

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right)$$

The degree of match between chunk i and the retrieval cues reflects the weight W associated with each retrieval cue j, which defaults to the total amount of goal activation G available divided by the number of cues (G/j). Weights are assumed to be equal across all cues. The degree of match between chunk i and the retrieval cues is the sum of the weighted associative boost for each retrieval cue S_j that matches a feature value of chunk i. The associative boost that a cue contributes to a matching chunk is reduced as a function of the "fan" of that cue, i.e., the number of competitor items in memory that also match the cue (Anderson 1974; Anderson & Reder 1999), according to Equation 3.

(8)   Equation 3

$$S_{ji} = S - \ln\left(\text{fan}_j\right)$$

Partial matching makes it possible to retrieve a chunk that matches only some of the cues (Anderson & Matessa 1997; Anderson et al. 2004), creating the opportunity for retrieval interference of the sort that leads to agreement attraction errors (e.g., Wagers et al. 2009), like those described in the Introduction of the current study. Partial matching is calculated as the matching summation over the k feature values of the retrieval cues. P is a match scale, and M_{ki} reflects the similarity between the retrieval cue value k and the value of the corresponding feature of chunk i, expressed by maximum similarity and maximum difference.

   Lastly, stochastic noise contributes to the activation level of chunk i. Noise is generated from a logistic distribution with a mean of 0, controlled by the noise parameter, which is related to the variance of the distribution, according to Equations 4 and 5. Noise is recomputed at each retrieval attempt.

---

[2] Equations 1–7 are based on ACT-R 6.0. Readers familiar with the (Lewis & Vasishth 2005) ACT-R model may notice the non-standard presentation of Equation 1: the sign on the partial match component has been moved outside of the summation to indicate its penalizing nature.

(9)   Equation 4

$$\epsilon \sim \text{logistic}\left(0, \sigma^2\right)$$

(10)   Equation 5

$$\sigma^2 = \frac{\pi^2}{3}s^2$$

Activation A_i determines the probability of retrieving a chunk according to Equation 6. The probability of retrieving chunk i is a logistic function of its activation with gain $1/s$ and threshold . Chunks with a higher activation are more likely to be retrieved.

(11)   Equation 6

$$P\left(\text{retrieval}\right) = \frac{1}{1 + e^{(-A_i - \tau)/s}}$$

Lastly, the mapping from activation to retrieval latency is calculated according to Equation 7, where T_i is the time it takes to retrieve an item from memory for further processing, and F is a scaling factor to ensure that model predictions are on an appropriate time scale. According to Equation 7, retrieval time is an inverse function of activation, such that a reduction in activation causes retrieval time to increase.

(12)   Equation 7

$$T_i = Fe_i^{-A_i}$$

To facilitate comparison with the acceptability data from Experiment 1, the current study adopts the standard linking assumption that there is a monotonic relationship between the retrieval latencies generated by the model and the behavioral measures that index retrieval operations, e.g., acceptability judgments (Anderson & Milson 1989; Anderson 1990; Lewis & Vasishth 2005; Vasishth et al. 2008; Kim et al. 2011; Dillon et al. 2013; Kush & Phillips 2014; Dillon et al. 2015; Parker & Lantz 2017). There are certainly additional factors beyond retrieval time that impact acceptability (e.g., structural attachment, discourse integration, ambiguity resolution, etc.), but it is assumed that they do not disrupt the monotonic relationship between retrieval latencies and acceptability. Importantly, previous studies on retrieval in sentence comprehension have shown that the retrieval latencies predicted by Equation 7 provide a good quantitative fit to acceptability data. (e.g., Lewis & Vasishth 2005; Vasishth et al. 2008).

3.1.2 Retrieval cues for ellipsis resolution

Following previous studies on antecedent-ellipsis mismatches (Arregui et al. 2006; Kim et al. 2011), the current model required a syntactically matching antecedent (i.e., syntactic identity), using the same cues for antecedent retrieval described in previous studies on ellipsis resolution (Martin & McElree 2008; 2009; 2011; Kim et al. 2011; Martin et al. 2012), including cues for matching syntactic category, e.g., VP, NP (Merchant 2013a), clause structure, e.g., main clause, embedded clause (Frazier & Clifton 2005; Frazier 2013), voice, e.g., active, passive, and morphological marking, such as "-en" for passive forms (Lasnik 1999; Kim et al. 2011). Other relevant cues described in the literature may include Merchant's E(llipsis)-feature, which is responsible for licensing ellipsis (see Merchant 2019 for a review). For current purposes, I assume the presence of an E-feature throughout, without further specification. A summary of all cue specifications for each test condition can be found in Table 3.

According to a retrieval-based account of ellipsis resolution, identity constraints at the ellipsis site provide instructions to the retrieval system to recover a matching antecedent with specific features, e.g., matching category, voice, and morphological structure, in a specific position, e.g., matching syntactic function, level of embedding, clause structure. This theory of cues is restrictive because it is bounded by the independently developed principles of the ACT-R framework (e.g., cues are derived from the current linguistic context and grammatical constraints and correspond to a subset of the features of the target), and the empirically motivated principle that there is a direct mapping between grammatical constraints at the dependency site and the cues used in retrieval (Van Dyke & McElree 2006; 2011; Parker & Phillips 2017).

3.1.3 Antecedent structures

The model assumes the same antecedent structures proposed in Kim et al. (2011) and Arregui et al. (2006), as shown in Figures 3 and 4. Following Kim et al. (2011), the voice head distinguishes active and passive antecedents. On their analysis, the passive voice head contributes an overt morpheme "-en", whereas the active voice head is phonologically null. This distinction is adopted in the current study, such that marking is not used
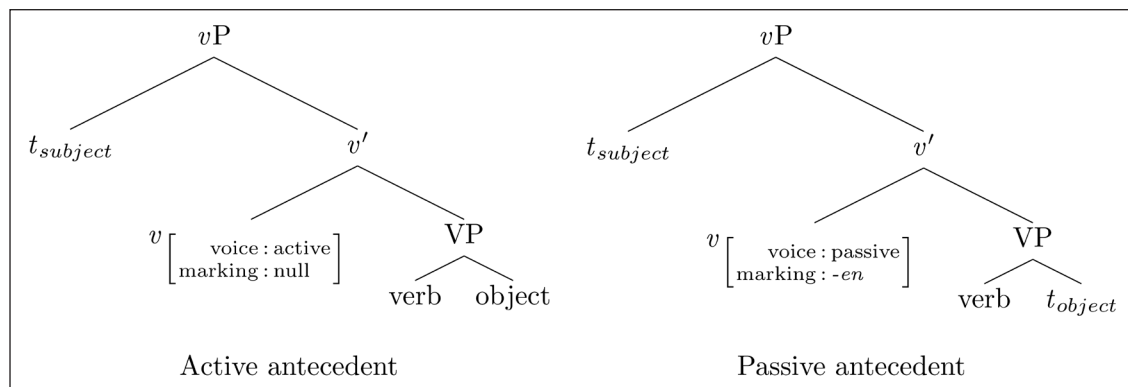

Figure 3: Assumed syntactic structures of active and passive antecedents (adapted from Kim et al. 2011).
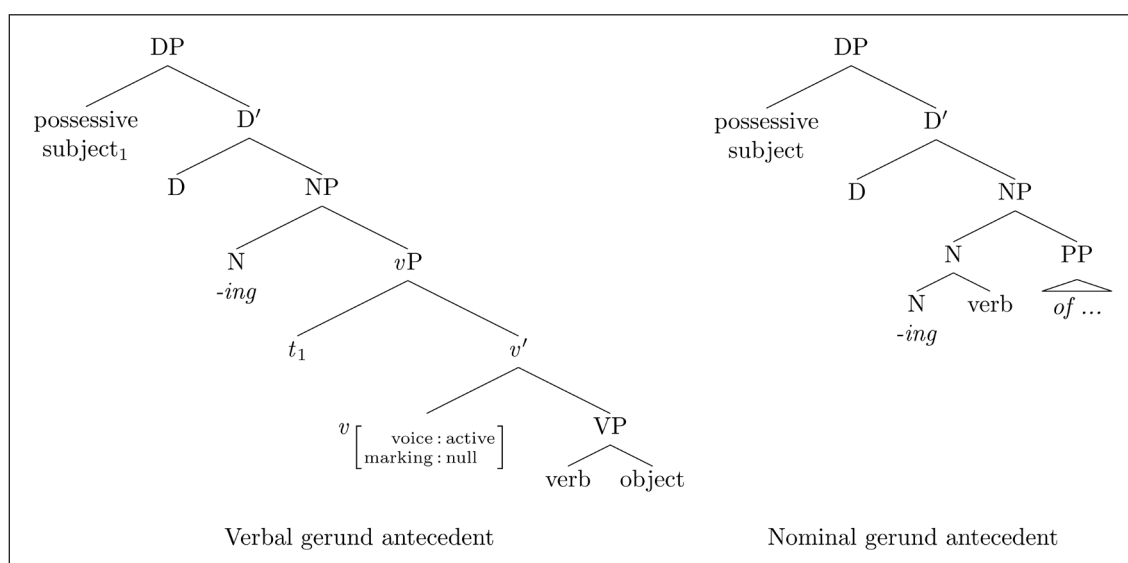

Figure 4: Assumed syntactic structures of verbal and nominal gerund antecedents (adapted from Arregui et al. 2006).

as a cue for active structures (represented in Table 3 as "null"). Following Arregui et al. (2006), the critical distinction between the gerund antecedents is that a VP is included in the verbal gerund, but not in the nominal gerund. Lastly, based on recent experimental evidence on ellipsis processing, it will be assumed that the ellipsis site includes a pointer that links the ellipsis clause to a syntactic representation of the antecedent in memory (Martin & McElree 2008; 2009; 2011). But the findings reported here are equally compatible with alternative accounts that assume structure sharing (Frazier & Clifton 2005) or a cost-free process that involves copying antecedent information into the ellipsis site (Frazier & Clifton 2001).

3.1.4 Mapping retrieval to acceptability

Acceptability follows from two properties of memory retrieval. The main driving force behind the model's predictions is the degree of match between the retrieval cues at the ellipsis site (described in §3.1.1) and the corresponding features of the antecedent structures encoded in memory (described in §3.1.2), such that more overlap increases activation of the antecedent (i.e., A_i in Equation 1), leading to a faster processing time (e.g., T_i in Equation 7) and boosted acceptability. The link between retrieval/processing time and acceptability is based on the independently motivated linking assumption shared by previous studies that there is a direct, inverse relationship between processing time (measured here in terms of retrieval latency) and acceptability (e.g., see Kim et al. 2011: fn. 13.), such that faster processing times correspond with higher acceptability, and slower processing times correspond with lower acceptability. A secondary factor that impacts processing time and hence acceptability, specifically for passive ellipsis structures, is an additional retrieval process required to interpret passive clauses. Passive clauses require retrieval of the subject NP at the object gap in the VP (Osterhout & Swinney 1993), independently of the retrieval for an antecedent. This additional retrieval will increase the overall time required to process and interpret the ellipsis clause, resulting in degraded acceptability relative to active structures, which do not require the additional retrieval.

To provide a concrete example, consider the cue specification for the (mis)matching forms in Table 2. For the Active-Active condition, the antecedent provides a perfect match to the retrieval cues, resulting in a high activation value for the antecedent, which facilitates retrieval and boosts its acceptability. By contrast, the antecedent in the Active-Passive condition mismatches several of the retrieval cues and requires an additional retrieval to resolve passivization, resulting in a relatively lower activation value, which slows retrieval and degrades acceptability.

Table 2: Example cue specifications for a voice match sentence and voice mismatch sentence.

Active-Active Voice Match: Jill betrayed Abby, and Matt did too.				
Retrieval cues	**Antecedent features**	**Cue match**	**Degree of match**	**Additional retrievals**
category: vP	vP	yes	full match	no
clause: main	main	yes		
voice: active	active	yes		
marking: null	null	null		
Active-Passive Voice Mismatch: Jill betrayed Abby, and Matt was too.				
Retrieval cues	**Antecedent features**	**Cue match**	**Degree of match**	**Additional retrievals**
category: vP	vP	yes	–2	yes
clause: main	main	yes		
voice: passive	active	no		
marking: -en	null	no		

A detailed summary of the retrieval cues and antecedent feature-value specifications for each condition in Table 1 is provided in Table 3. Table 3 also specifies the overall degree of probe-to-target match and whether an additional retrieval at the VP object position is required (i.e., for passive ellipsis clauses) to contextualize the predicted differences in acceptability.

3.2 Simulations

5,000 Monte Carlo simulations were run for each condition in Table 1 (i.e., Active-Active, Passive-Passive, Active-Passive, Passive-Active, Verbal Gerund, and Nominal Gerund). All parameters were set to the default values described in previous work (Lewis & Vasishth 2005; Vasishth et al. 2008), with the exception of the Latency Factor (F), which was chosen to set the predictions on the appropriate time scale for sentence processing. The same parameter setting was applied to each condition. A comprehensive list of the parameter values used in the current study is shown in Table 4. Each simulation included the full series of hypothesized retrievals, based on Equations 1–7. I report the average predicted latency by condition, which provides a measure of how long on average antecedent retrieval took in each condition.

Table 3: Retrieval cues and antecedent features for all conditions.

Retrieval cues	Antecedent features	Cue match	Degree of match	Additional retrievals
Active-Active Voice Match				
category: vP clause: main voice: active marking: null	vP main active null	yes yes yes null	full match	no
Passive-Passive Voice Match				
category: vP clause: main voice: passive marking: -en	vP main passive -en	yes yes yes yes	full match	yes
Verbal Gerund Mismatch				
category: vP clause: main voice: active marking: null	vP embedded active null	yes no yes null	–1	no
Nominal Gerund Mismatch				
category: vP clause: main voice: active marking: null	NP embedded active null	no no yes null	–2	no
Passive-Active Voice Mismatch				
category: vP clause: main voice: active marking: null	vP main passive -en	yes yes no null	–1	yes
Active-Passive Voice Mismatch				
category: vP clause: main voice: passive marking: -en	vP main active null	yes yes no no	–2	yes

Table 4: Parameter settings for the computational model.

Parameter	Value
Latency Factor (F)	0.20
Decay (d)	0.50
Total source activation (G)	1.00
Maximum associative strength (S)	1.50
Maximum difference (P)	-0.60
Noise ()	0.45

3.3 Model results

Figure 5 shows the model's predicted mean retrieval latencies for each condition in Table 1, mapped to the corresponding acceptability judgments from Experiment 1. Across conditions, there is a tightly coupled inverse relation between predicted retrieval latencies and acceptability, such that as retrieval latencies increase, acceptability decreases. Overall, the model accurately predicted the three key components of the acceptability profile: matching voice > gerundive antecedents > mismatching voice (where > indicates "more acceptable than"). For voice match sentences (Active-Active vs. Passive-Passive), the model predicted increased difficulty for passive-passive sentences, which maps to decreased acceptability, due to the increased processing time associated with passive clause structures, i.e., retrieval of the subject noun phrase at the object gap following the verb. For the voice mismatch sentences (Passive-Active vs. Active-Passive), the model predicted increased difficulty for Active-Passive forms. Both mismatches involve additional routine retrievals to resolve passivization, and hence increased processing time, which maps to decreased acceptability relative to voice matches. But Active-Passive mismatches are predicted to be more difficult than the Passive-Active mismatches due to the combination of mismatches on voice and morphological cues, e.g., passive + -en, as shown in Table 3. For gerundive antecedents, the model predicted increased difficulty for nominal gerunds. In the case of the verbal gerund mismatch, a VP antecedent is available, but in an unexpected structural position, yielding a mismatch on the clause cue. By contrast, no suitable VP antecedent exists for nominal gerunds, as noted by Arregui et al. (2006), violating position and category expectations, as shown in Table 3. Taken together, these results can be considered a success for the model, as there is a direct correspondence between predicted processing time and acceptability (adjusted $R^2 = 0.88$), without adjusting the model parameters.

4 General discussion

4.1 Summary of results

The present study used computational modeling to better understand why antecedent-ellipsis mismatches vary in acceptability. Previously, antecedent-ellipsis mismatches motivated the proposal that ellipsis resolution relies on special processing rules that either (i) repair and restructure an ungrammatical antecedent to satisfy syntactic identity constraints at the ellipsis site (Arregui et al. 2006), or (ii) over-generate a set of potential antecedent representations that are ranked and searched based on grammar-independent processing heuristics (Kim et al. 2011). These accounts reinforce the standard view in (psycho)linguistics that the parser and grammar reflect distinct structure-building systems that operate on different time scales using a distinct set of rules and representations (Fodor et al. 1974; Levelt 1974; Townsend & Bever 2001; Ferreira et al. 2002; Ferreira & Patson 2007).
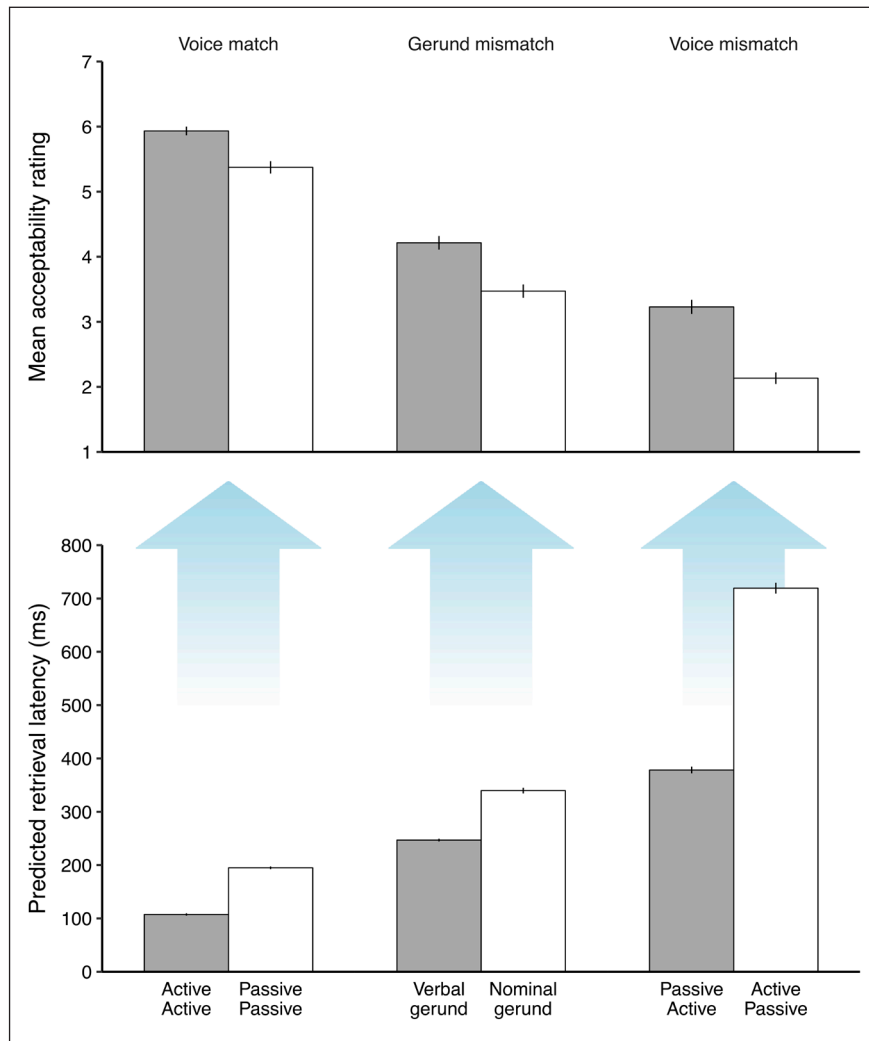
Figure 5: The antecedent-ellipsis acceptability cline presented as predicted retrieval latency for the antecedent-ellipsis (mis)matches in Table 1, mapped to the corresponding judgments from Experiment 1.

The alternative view pursued in the current study is that antecedent-ellipsis mismatches are the product of a single structure-building system (i.e., the grammar) implemented in a noisy cognitive architecture. This hypothesis was motivated by a series of recent studies on a related class of acceptable ungrammaticalities involving subject-verb agreement, reflexive licensing, case licensing, and negative polarity item licensing, which claim that spurious acceptability arises not because of the application of parser-specific heuristics, repair rules, or a parser-grammar misalignment, but rather because the constraints on linguistic dependency formation are implemented using a noisy memory retrieval system that creates the opportunity for errors (Vasishth et al. 2008; Wagers et al. 2009; Phillips, et al. 2011; Dillon et al. 2013; Sloggett 2013; Tanner et al. 2014; Lewis & Phillips 2015; Tucker et al. 2015; Parker & Phillips 2016; 2017; Parker & Lantz 2017). Based on these findings, the present study tested whether antecedent-ellipsis mismatches could be accounted for in a similar fashion, leading to a more unified account of acceptable ungrammaticalities.

To this end, Experiment 1 used untimed acceptability judgments to confirm the acceptability cline reported in previous studies (Arregui et al. 2006; Kim et al. 2011). Results replicated each portion of the acceptability cline reported in the literature, based on a within-participant comparison of the key mismatches using a single, uniform acceptability measure across ellipsis forms. Then, as proof-of-concept, Experiment 2 used an established computational model of memory retrieval (ACT-R; Lewis & Vasishth 2005) to

show how the observed acceptability profile follows directly from independently motivated principles of working memory without invoking multiple representational systems. Specifically, the model predicted differences in processing time as a function of the degree of match between the retrieval cues at the ellipsis site and the features of the antecedent (i.e., the independently motivated principle of probe-to-target similarity), with a tightly coupled inverse relationship between processing time and acceptability.

4.2　What distinguishes the retrieval-based account?

All accounts of the VP ellipsis acceptability cline, including the current account, accurately predict certain aspects of the acceptability profile, and all accounts link acceptability to the extent of the syntactic mismatch. So, what distinguishes the current retrieval-based account from previous accounts (e.g., Arregui et al. 2006; Kim et al. 2011; Kertz 2013)? I suggest that there are three distinguishing factors.

First, the current retrieval-based account provides an explicit description of how ellipsis is resolved during moment-by-moment processing and how online processes map to acceptability judgments. Previous accounts (i.e., Arregui et al. 2006; Kim et al. 2011; Kertz, 2013) are underspecified about the precise moment-by-moment parser operations, and do not provide sufficient details about the cost metrics associated with the proposed processing operations (i.e., repair, information structure analysis, search) or how they are applied in real time. For instance, Kim and colleagues reference parsing strategies involving a serial search procedure, but they do not describe the costs associated with a serial search. Furthermore, the proposal for a serial search predicts response times that are incompatible with the empirical data on retrieval times in sentence comprehension (e.g., Murdock 1971; McElree 2000; 2006; McElree et al. 2003). The current retrieval account, by contrast, explicitly defines each step of the computation, and provides a detailed description of the associated costs and how they map to behavioral measures, based on independently motivated principles of working memory. However, it is important to emphasize that there are limitations on the inferences that we can draw from the processing of ungrammatical strings. An important task for future work is to explore the predictions of the current retrieval-based account under a broader range of grammatical configurations. For instance, one prediction of the current account is that introducing multiple VPs that match the proposed retrieval cues should trigger similarity-based interference, in both grammatical and ungrammatical configurations. This prediction will be tested in future research.

Second, all accounts must consider that real-time ellipsis resolution relies on memory retrieval mechanisms to access the antecedent (Martin & McElree 2008; 2009; 2011; Martin et al. 2012), and that retrieval processes directly impact acceptability judgments (Bever 1970; Dillon et al. 2015; Kim & Runner 2018). If key portions of the acceptability cline can be explained based solely on the routine operations of the memory retrieval system, as shown in the current study, then the motivation for additional extra-grammatical operations is reduced. Importantly, the current retrieval-based account does not deny that additional post-retrieval factors (e.g., information structure constraints, antecedent reanalysis) can further impact acceptability, but rather claims that post-retrieval operations cannot be the source of the observed effects. To elaborate, retrieval is necessary, but not sufficient for interpretation, and ellipsis resolution does not culminate upon completion of memory access. As such, there are many additional constraints on interpretation, such as those involving information-structural constraints (e.g., Kertz 2013) and reanalysis (e.g., Arregui et al. 2006) that may be applied as well-formedness checks on the output of retrieval, further shaping acceptability. The proposal that certain constraints are applied as post-retrieval checks on dependency resolution is not new (e.g., Dillon et al. 2013 for

a similar proposal for reflexive licensing), and the application of those constraints will feed interpretation and acceptability as comprehension unfolds over time. However, these additional processes should not disrupt the monotonic relationship between the retrieval operations that guide initial processing and acceptability. It is clear from the good model fits obtained in the current study that the memory system places a strong constraint on the acceptability profile.

The third factor that distinguishes the current retrieval-based account from previous accounts is that the current account improves empirical coverage. Specifically, the retrieval-based account captures four facts that existing accounts fail to explain. These differences are summarized below.

1. **Active-Active > Passive-Passive:** Existing accounts do not explain why Active-Active matches are more acceptable than Passive-Passive matches. More specifically, existing accounts do not predict a difference between these forms. According to previous accounts, acceptability for active-active and passive-passive matches should be identical, since both involve a fully-matched antecedent. Under the retrieval-based account, the degraded acceptability of passive-passive matches reflects increased processing cost due to additional retrievals to resolve passivization (see § 3.1.3). The current account provides an explicit description of these processes.

2. **Passive-Active > Active-Passive:** Existing accounts predict no difference between voice mismatches. Under the repair account proposed by Arregui et al. (2006), restructuring a passive antecedent to create an active match and vice versa should involve the same number of repair steps. Arregui and colleagues argue that there should be a difference between these mismatches, based on early findings that comprehenders are more likely mis-recall an active as a passive than the other way around (Mehler 1963). But in the previous studies on ellipsis mismatches, participants were not asked for an explicit recall judgment or required to recall the sentence verbatim, as in Mehler's study, and there is little reason to believe that comprehenders incorrectly encoded or misremembered the passive structures used in the current materials (which were obtained from Arregui et al. 2006). The search-based account proposed by Kim et al. (2011) assumes that ellipsis for both mismatches targets the same projection, in which case both mismatches should involve the same amount of search work to recover a matching antecedent. Under the retrieval-based account, the differences in acceptability between these mismatches reflect more mismatching cues required by the passive ellipsis.

3. **Voice matches > Verbal gerund mismatches:** Kim et al. (2011: 349) noted that it is unclear under existing accounts why verbal gerunds are less acceptable than normal VPs, since a VP antecedent is available in the syntactic representation of the sentence. The current account offers an explanation for this difference. Under the retrieval-based account, verbal gerunds are less acceptable because they violate the expectation for an antecedent which occurs as a matrix, as opposed to embedded, verb phrase. This effect may be grounded in Frazier and Clifton's "main assertion principle", which favors antecedents in the main clause (Frazier & Clifton 2005; Frazier 2013). This preference is implemented in the cue structure with the clause cue.

4. **Gerundive antecedent mismatches > Voice mismatches:** Existing accounts do not offer an explanation for why voice mismatches are worse than gerundive mismatches. Under the current retrieval-based account, voice

mismatches involve mismatching cues and additional retrievals (and hence increased processing costs) to resolve passivization. Likewise, the gerund constructions also contain a main clause VP that matches some of the retrieval cues, e.g., [+VP, +main clause], which helps boost their acceptability. But this VP does not bear the appropriate E-feature required to license ellipsis (Merchant 2011).

4.3 Questions for future research

The present study showed that key aspects of the ellipsis acceptability cline can be captured with a single structure-building system implemented in a noisy memory retrieval system. These findings are consistent with recent arguments that the acceptable ungrammaticalities observed for other linguistic dependencies are the product of noisy memory retrieval mechanisms (Vasishth et al. 2008; Wagers et al. 2009; Phillips et al. 2011; Dillon et al. 2013; Sloggett 2013; Tanner et al. 2014; Lewis & Phillips 2015; Tucker et al. 2015; Parker & Lantz 2017; Parker & Phillips 2016; 2017). The current results thus contribute to a more unified account of acceptable ungrammaticalities across linguistic dependencies and provide a theoretical foundation to investigate other types of antecedent-ellipsis mismatches discussed in the literature.

For instance, an important task for future research is to determine how to capture the "comet" mismatches described in Arregui et al. (2006), e.g., Seeing the comet was nearly impossible, but John did see the comet, and the category mismatches discussed in Kim et al. (2011), e.g., An admission of guilt was needed, but the suspect wouldn't admit guilt. It is possible that the comet mismatches could be explained based on an extension of the current proposal for the gerundive mismatches by exploiting the expectation for a matrix, rather than an embedded, VP (Frazier & Clifton 2005; Frazier 2013). Likewise, category mismatches like those tested by Kim et al. (2011) could be explained by exploiting the expectation for a VP antecedent using the corresponding category cues. I leave further investigation of this suggestion to future work.

Another type of mismatch worth investigating under the current proposal includes the so-called "tolerable" vs. "intolerable" mismatches, e.g., John likes this movie and Bill might like this movie too, vs. *John is fond of this movie and Bill might be fond of this movie too (Lipták 2015). It is unclear at present how the acceptability profile of tolerable vs. intolerable mismatches might be captured under the current retrieval-based framework, without appealing to post-retrieval semantic or discourse-level processes. A similar problem arises for the contrast between voice mismatches in sentences like *This problem was looked into by John, and Bob did look into this problem too, vs. This problem was to have been looked into, but obviously nobody did look into this problem (Kehler 1993b). These sentences are problematic for existing accounts because both involve a passive-active mismatch but differ in acceptability. One possibility is that this contrast might involve an interaction of memory retrieval effects, syntactic processing, and discourse-level processing, as discussed in recent accounts of ellipsis resolution (Kim & Runner 2018). An important task for future work is to tease apart the contributions of these different factors.

Another question is how the current retrieval-based account applies to other types of ellipsis, such as sluicing (IP ellipsis). Both VP ellipsis and sluicing rely on the same memory retrieval system (Martin & McElree 2011), but unlike VP ellipsis, sluicing does not tolerate voice mismatches (Merchant 2001; 2013b). One potential explanation for this difference is that sluicing and VP ellipsis rely on different retrieval cues. For instance, VP-ellipsis relies on a VP category cue with corresponding verbal cues (e.g., voice, morphology) to

guide retrieval, but sluicing relies on cues at the IP level, and thus does not deploy the verbal cues that give rise to gradient acceptability for mismatching VPs. At this level of analysis, the different behaviors for sluicing vs. VP ellipsis with respect to voice mismatch effects might reflect differences in what cues the grammar makes available to the retrieval system.

Lastly, it is worth discussing how the current retrieval-based approach fits with existing formal theories of VP-ellipsis. Formal theories of ellipsis typically fall into one of two categories. Syntactic theories assume that the content of the ellipsis site involves detailed structure, whereas referential theories assume that the ellipsis site involves a null proform/pointer, akin to other types of referential expressions, such as pronouns (e.g., Tanenhaus & Carlson 1990; Hardt 1993; Ginzburg & Sag 2000; Culicover & Jackendoff 2005; Martin & McElree 2008). Some syntactic theories assume that a full copy of the antecedent is reconstructed at the ellipsis site, whereas referential theories assume that the ellipsis site does not include a copy of the antecedent, but rather a direct link to the antecedent (see Phillips & Parker 2014 for a review). The current retrieval-based account fits most naturally with a pointer-style analysis of ellipsis resolution, which is consistent with recent experimental work showing that real-time ellipsis resolution is mediated by a pointer mechanism (e.g., Martin & McElree 2008). However, the current account does not rule out the possibility that a representation of the antecedent is reconstructed at the ellipsis site, possibly as a post-retrieval operation.

5 Conclusion

Previously, acceptable but ungrammatical cases of VP-ellipsis have been presented as evidence for a special class of extra-grammatical processing heuristics that reinforce the standard grammar-parser distinction. The current study provides evidence from computational modeling that it is not necessary to posit special, processor-specific heuristics or a parser-grammar distinction to capture the ellipsis acceptability cline. Rather, the observed profiles can be explained using a single structure-building system that relies on noisy memory retrieval mechanisms to implement language-specific tasks, like dependency formation. Importantly, the current retrieval-based account improves empirical coverage. Broadly, these results provide new insights into how ellipsis structures are processed in real time and offer a more unified account of how processing models relate to formal accounts of grammatical phenomena.

Additional File

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

Acknowledgements

Competing Interests

The author has no competing interests to declare.

References

Anderson, John Robert. 1974. Retrieval of propositional information from long-term memory. *Cognitive Psychology* 6. 451–474. DOI: https://doi.org/10.1016/0010-0285(74)90021-8

Anderson, John Robert. 1990. *The adaptive character of thought*. Hillsdale, NJ: Erlbaum. DOI: https://doi.org/10.4324/9780203771730

Anderson, John Robert, Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere & Yulin Qin. 2004. An integrated theory of the mind. *Psychology Review* 111. 1036–1060. DOI: https://doi.org/10.1037/0033-295X.111.4.1036

Anderson, John Robert & Lynne M. Reder. 1999. The fan effect: New results and new theories. *Journal of Experimental Psychology: General* 128. 186–197. DOI: https://doi.org/10.1037/0096-3445.128.2.186

Anderson, John Robert & Michael Matessa. 1997. A production system theory of serial memory. *Psychological Review* 104. 728–748. DOI: https://doi.org/10.1037/0033-295x.104.4.728

Anderson, John Robert & Robert Milson. 1989. Human memory: An adaptive perspective. *Psychology Review* 96. 703–719. DOI: https://doi.org/10.1037/0033-295x.96.4.703

Arregui, Ana, Charles Clifton, Jr., Lyn Frazier & Keir Moulton. 2006. Processing elided VPs with flawed antecedents. *Journal of Memory and Language* 55. 232–246. DOI: https://doi.org/10.1016/j.jml.2006.02.005

Baayen, R. H., D. J. Davidson & D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59. 390–412. DOI: https://doi.org/10.1016/j.jml.2007.12.005

Badecker, William & Richard L. Lewis. 2007. A new theory and computational model of working memory in sentence production: Agreement errors as failures of cue-based retrieval. *Talk at the 20th CUNY Conference on Human Sentence Processing*. University of California, San Diego.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278. DOI: https://doi.org/10.31234/osf.io/39mhs

Bates, D., Martin Maechler, Ben Bolker & Steven Walker. 2018. *lme4: Linear mixed-effects models using S4 classes*. Retrieved from: http://cran.r-project.org/package=lme4.

Berman, Marc G., John Jonides & Richard L. Lewis. 2009. In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35. 317–353. DOI: https://doi.org/10.1037/Fa0014873

Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199544004.013.0030

Dalrymple, Mary, Stuart M. Shieber & Fernando C. N. Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy* 14. 399–452. DOI: https://doi.org/10.1007/bf00630923

Dillon, Brian, Adrian Staub, Joshua Levy & Charles Clifton, Jr. 2015. Distinguishing discrete from gradient grammaticality using Likert scale data. *Poster at the 28th CUNY Human Sentence Processing Conference*, University of Southern California.

Dillon, Brian, Alan Mishler, Shayne Sloggett & Colin Phillips. 2013. Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language* 69. 85–103. DOI: https://doi.org/10.1016/j.jml.2013.04.003

Ferreira, Fernanda, Karl D. G. Bailey & Vittoria Ferraro. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11. 11–15. DOI: https://doi.org/10.1111/1467-8721.00158

Ferreira, Fernanda & Nikole D. Patson. 2007. The 'good enough' approach to language comprehension. *Language and Linguistics Compass* 1. 71–83. DOI: https://doi.org/10.1111/j.1749-818x.2007.00007.x

Fiengo, Robert & Robert May. 1994. *Indicies and identity*. Cambridge, MA: MIT Press. DOI: https://doi.org/10.1017/s0022226700000840

Fodor, Jerry A., Thomas G. Bever & Merrill F. Garrett. 1974. *The psychology of language*. New York: McGraw-Hill. DOI: https://doi.org/10.2307/412725

Frazier, Lyn. 2013. A recycling approach to processing ellipsis. In Lisa Lai-Shen Cheng & Nobert Corver (eds.), *Diagnosing syntax*, 485–501. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199602490.001.0001

Frazier, Lyn & Charles Clifton, Jr. 2001. Parsing coordinates and ellipsis: Copy α. *Syntax* 4. 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. 421–430. DOI: https://doi.org/10.1111/j.1467-9612.2005.00077.x

Gelman, Andrew & Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/cbo9780511790942

Ginzburg, Jonathan & Ivan Sag. 2000. *Interrogative investigations*. Stanford, CA: CSLI Publications.

Hardt, Daniel. 1992. VP-ellipsis and contextual interpretation. In *COLING '92 Proceedings of the 14th conference on computational linguistics* 1. 303–309. DOI: https://doi.org/10.3115/992066

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

Hardt, Daniel. 1999a. Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy* 22. 187–221. DOI: https://doi.org/10.1023/A:1005427813846

Hardt, Daniel. 1999b. VPE as a proform: Some consequences for binding. In Francis Corblin, Jean-Marie Marandin & Daniele Godard (eds.), *Empirical issues in formal syntax and semantics 2: Selected papers from the Colloque de Syntaxe et Sèmantique à Paris*, 215–232. The Hague: Thesus. DOI: https://doi.org/10.1075/sl.23.1.13bou

Kehler, Andrew. 1993a. A discourse copying algorithm for ellipsis and anaphora resolution. *Paper presented at the 31st Conference of the European Chapter of the Association for Computational Linguistics (EACL-93)*. Utrecht, the Netherlands. DOI: https://doi.org/10.3115/976744.976769

Kehler, Andrew. 1993b. The effect of establishing coherence in ellipsis an anaphora resolution. *Paper presented at the 31st Conference of the Association for Computational Linguistics (ACL-93)*, Columbus, OH. DOI: https://doi.org/10.3115/981574.981583

Kehler, Andrew. 2000. Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23. 533–575. DOI: https://doi.org/10.1023/A:1005677819813

Kertz, Laura. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89. 390–428. DOI: https://doi.org/10.1353/lan.2013.0051

Kim, Christina S., Gregory M. Kobele, Jeffrey T. Runner & John T. Hale. 2011. The acceptability cline in VP-ellipsis. *Syntax* 14. 318–354. DOI: https://doi.org/10.1111/j.1467-9612.2011.00160.x

Kim, Christina S. & Jeffrey T. Runner. 2018. The division of labor in explanations of verb phrase ellipsis. *Linguistics and Philosophy* 41. 41–85. DOI: https://doi.org/10.1007/s10988-017-9220-0

Kohonen, Tuevo. 1980. *Content-addressable memories*. Berlin, DE: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-642-96552-4

Kush, Dave & Colin Phillips. 2014. Local anaphor licensing in an SOV language: Implications for retrieval strategies. *Frontiers in Psychology* 5. 1–12. DOI: https://doi.org/10.3389/fpsyg.2014.01252

Lago, Sol, Diego E. Shalom, Mariano Sigman, Ellen F. Lau & Colin Phillips. 2015. Agreement processes in Spanish comprehension. *Journal of Memory and Language* 82. 133–149. DOI: https://doi.org/10.31219/osf.io/5rm3z

Lappin, Shalom. 1992. The syntactic basis of ellipsis resolution. In Steve Berman & Arild Hesttvik (eds.), *Proceedings of the Stuttgart Ellipsis Workshop*: *Working papers of the Sonderforschungsbereich 340, Report No. 29-1993, SFB, 340*. University of Tübingen.

Lappin, Shalom. 1996. The interpretation of ellipsis. In Shalom Lappin (ed.), *The handbook of contemporary semantic theory*, 145–175. Oxford, UK: Blackwell. DOI: https://doi.org/10.1111/b.9780631207498.1997.00003.x

Lascarides, Alex & Nicholas Asher. 1993. Temporal interpretation, discourse relations and common sense entailment. *Linguistics and Philosophy* 16. 437–493. DOI: https://doi.org/10.1007/bf00986208

Lasnik, Howard. 1999. Verbal morphology: Syntactic structures meets the minimalist program. In Howard Lasnik (ed.), *Minimalist analysis,* 97–119. Malden, MA: Blackwell.

Levelt, W. J. M. 1974. *Formal grammars and psycholinguistics*. The Hague: Mouton. DOI: https://doi.org/10.1515/9783110889390

Lewandowsky, Stephan, Klaus Oberauer & Gordon D. A. Brown. 2009. No temporal decay in verbal short-term memory. *Trends in Cognitive Science* 13. 120–126. DOI: https://doi.org/10.1016/j.tics.2008.12.003

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

Lewis, Richard L., Shravan Vasishth & Julie A. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10. 447–454. DOI: https://doi.org/10.1016/j.tics.2006.08.007

Lewis, Shevaun & Colin Phillips. 2015. Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research* 44. 27–46. DOI: https://doi.org/10.1007/s10936-014-9329-z

Lipták, Anikó. 2015. Identity in ellipsis: An introduction. *Lingua* 166. 155–171. DOI: https://doi.org/10.1016/j.lingua.2015.08.003

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

Martin, Andrea E. & Brian McElree. 2009. Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35. 1231–1239. DOI: https://doi.org/10.1037/a0016271

Martin, Andrea E. & Brian McElree. 2011. Direct-access retrieval during sentence comprehension: Evidence from Sluicing. *Journal of Memory and Language* 64. 327–343. DOI: https://doi.org/10.1016/j.jml.2010.12.006

Martin, Andrea E., Mante S. Nieuwland & Manuel Carreiras. 2012. Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage* 59. 1859–1869. DOI: https://doi.org/10.1016/j.neuroimage.2011.08.057

Martin, Andrea E., Mante S. Nieuwland & Manuel Carreiras. 2014. Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain & Language* 135. 42–51. DOI: https://doi.org/10.1016/j.bandl.2014.05.001

McElree, Brian. 2000. Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research* 29. 155–200. DOI: https://doi.org/10.1023/A:1005184709695

McElree, Brian. 2006. Accessing recent events. In Brian H. Ross (ed.), *The psychology of learning and motivation – Advances in research and theory*, 155–200. San Diego, CA: Academic Press.

McElree, Brian & Barbara Anne Dosher. 1989. Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology* 18. 346–373. DOI: https://doi.org/10.1037//0096-3445.118.4.346

McElree, Brian & Barbara Anne Dosher. 1993. Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General* 122. 291–315. DOI: https://doi.org/10.1037/0096-3445.122.3.291

McElree, Brian, Stephanie Foraker & Lizbeth Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48. 67–91. DOI: https://doi.org/10.1016/s0749-596x(02)00515-6

Mehler, Jacques. 1963. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior* 2. 346–351. DOI: https://doi.org/10.1016/s0022-5371(63)80103-6

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

Merchant, Jason. 2013a. Polarity items under ellipsis In Lisa Lai-Shen Cheng & Nobert Corver (eds.), *Diagnosing syntax*, 441–462. Oxford, UK: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199602490.003.0022

Merchant, Jason. 2013b. Voice and ellipsis. *Linguistic Inquiry* 44. 77–108. DOI: https://doi.org/10.1162/ling_a_00120

Merchant, Jason. 2019. Ellipsis: A survey of analytical approaches. In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *Handbook of ellipsis*. Oxford, UK: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195375640.003.0001

Miller, George A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 101. 343–352. DOI: https://doi.org/10.1037/h0043158

Murdock, Bennet B. 1971. A parallel-processing model for scanning. *Perception & Psycophysics* 10. 289–291. DOI: https://doi.org/10.3758/bf03212827

Nairne, James S. 2002a. The myth of the encoding-retrieval match. *Memory* 10. 389–395. DOI: https://doi.org/10.1080/09658210244000216

Nairne, James S. 2002b. Remembering over the short-term: The case against the standard model. *Annual Review of Psychology* 53. 53–81. DOI: https://doi.org/10.1146/annurev.psych.53.100901.135131

Osterhout, Lee & David A. Swinney. 1993. On the temporal course of gap-filling during comprehension of verbal passives. *Journal of Psycholinguistic Research* 22. 273–286. DOI: https://doi.org/10.1007/BF01067834

Paape, Dario. 2016. Filling the silence: Reactivation, not reconstruction. *Frontiers in Psychology* 7. 1–18. DOI: https://doi.org/10.3389/fpsyg.2016.00027

Parker, Dan & Colin Phillips. 2016. Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition* 157. 321–339. DOI: https://doi.org/10.1016/j.cognition.2016.08.016

Parker, Dan & Colin Phillips. 2017. Reflexive attraction in comprehension is selective. *Journal of Memory and Language* 94. 272–290. DOI: https://doi.org/10.1016/j.jml.2017.01.002

Parker, Dan & Daniel Lantz. 2017. Encoding and accessing linguistic representations in a dynamically structured holographic memory system. *Topics in Cognitive Science* 9. 51–68. DOI: https://doi.org/10.1111/tops.12246

Parker, Dan, Michael Shvartsman & Julie A. Van Dyke. 2017. The cue-based retrieval theory of sentence comprehension: New findings and new challenges. In Linda Escobar, Vincent Torrens & Teresa Parodi (eds.), *Language processing and disorders,* 121–144. Newcastle, UK: Cambridge Scholars Publishing.

Patil, Umesh, Sharvan Vasishth & Richard L. Lewis. 2016. Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology* 7. 1–18. DOI: https://doi.org/10.3389/fpsyg.2016.00329

Pearlmutter, Neal J., Susan M. Garnsey & Kathryn Bock. 1999. Agreement processes in sentence comprehension. *Journal of Memory and Language* 41. 427–456. DOI: https://doi.org/10.1006/jmla.1999.2653

Peterson, Lloyd & Margaret Jean Peterson. 1959. Short-term retention of invidual verbal items. *Journal of Experimental Psychology* 58. 193–198. DOI: https://doi.org/10.1037/h0049234

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

Phillips, Colin, Matthew W. Wagers & Ellen F. Lau. 2011. Grammatical illusions and selective fallibility in real-time language comprehension. In Jeffrey T. Runner (ed.), *Experiments at the interfaces* 37. 147–180. Bingley, UK: Emerald Publications. DOI: https://doi.org/10.1163/9781780523750_006

Pollard, Carl & Ivan Sag. 1994. *Head-driven phrase structure grammar*. Chicago, IL: University of Chicago Press.

R Development Core Team. 2014. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: http://www.r-project.org/.

Sag, Ivan. 1976. *Deletion and logical form*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Sloggett, Shayne. 2013. Case licensing in processing: Evidence from German. *Poster at the 26th CUNY Conference on Human Sentence Processing*. University of South Carolina.

Sternberg, Saul. 1975. Memory-scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology* 27. 1–32. DOI: https://doi.org/10.1080/14640747508400459

Takahashi, Shoichi & Danny Fox. 2005. MaxElide and the re-binding problem. *Paper presented at SALT 15*. Ithica, NY. DOI: https://doi.org/10.3765/salt.v15i0.3095

Tanenhaus, Michael K. & Greg N. Carlson. 1990. Comprehension of deep and surface verb phrase anaphora. *Language and Cognitive Processes* 5. 257–280. DOI: https://doi.org/10.1080/01690969008407064

Tanner, Darren, Janet Nicol & Laurel Brehm. 2014. The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language* 76. 195–215. DOI: https://doi.org/10.1016/j.jml.2014.07.003

Townsend, David J. & Thomas G. Bever. 2001. *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Tucker, Matthew A., Ali Idrissi & Diogo Almeida. 2015. Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in Psychology* 6. 1–21. DOI: https://doi.org/10.3389/fpsyg.2015.00347

van Craenenbroeck, Jeroen. 2010. *The syntax of ellipsis: Evidence from Dutch dialects*. Oxford, UK: Oxford University Press.

Van Dyke, Julie A. & Brian McElree. 2006. Retrieval interference in sentence comprehension. *Journal of Memory and Language* 55. 157–166. DOI: https://doi.org/10.1016/j.jml.2006.03.007

Van Dyke, Julie A. & Brian McElree. 2011. Cue-dependent interference in comprehension. *Journal of Memory and Language* 65. 247–263. DOI: https://doi.org/10.1016/j.jml.2011.05.002

Vasishth, Shravan, Sven Brüssow, Richard L. Lewis & Heiner Drenhaus. 2008. Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science* 32. 685–712. DOI: https://doi.org/10.1080/03640210802066865

Wagers, Matthew W., Ellen F. Lau & Colin Phillips. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61. 206–237. DOI: https://doi.org/10.1016/j.jml.2009.04.002

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