The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval

Nayoun Kim, Laurel Brehm & Masaya Yoshida

Routledge
Taylor & Francis Group

Check for updates

REGULAR ARTICLE

The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval

Nayoun Kim[a], Laurel Brehm[b] and Masaya Yoshida[a]

[a]Department of Linguistics, Northwestern University, Evanston, IL, USA; [b]Max Planck Institute for Psycholinguistics, Max-Planck-Gesellschaft, Nijmegen, The Netherlands

ABSTRACT

We investigate whether grammatical information is accessed in processing noun phrase ellipsis (NPE) and other anaphoric constructions. The first experiment used an agreement attraction paradigm to reveal that ungrammatical plural verbs following NPE with an antecedent containing a plural modifier (e.g. *Derek's key to the boxes … and Mary's_ probably *are safe in the drawer*) show similar facilitation to non-elided NPs. The second experiment used the same paradigm to examine a coordination construction without anaphoric elements, and the third examined anaphoric *one*. Agreement attraction was not observed in either experiment, suggesting that processing NPE is different from processing non-anaphoric coordination constructions or anaphoric *one*. Taken together, the results indicate that the parser is sensitive to grammatical distinctions at the ellipsis site where it prioritises and retrieves the head at the initial stage of processing and retrieves the local noun within the modifier phrase only when it is necessary in parsing NPE.

1. Introduction

Successful real-time sentence processing requires establishing dependencies. For example, in English, the subject noun phrase (NP) controls agreement morphology on the verb, as illustrated in (1).

(1) a. He is in the room.
    b. *He are in the room.
    c. *They is in the room.
    d. They are in the room.

When the subject is singular, the verb must take a singular inflection, and when the subject is plural, the verb must take a plural inflection. This means that the number morphology of the verb is dependent upon the number of the subject noun.

This illustrates a broader principle: during online processing of a sentence involving a dependency relation, the parser needs to link the dependent element to its controlling element. It is often the case that the dependent element, which signals the presence of a dependency relation, is located after the controlling element. This means that when the dependent element is encountered, the parser must recognise a dependency relation and trigger the retrieval of a controlling element from memory in order to achieve the correct interpretation of a sentence.

In this series of experiments, we study subject-Aux agreement in the context of elided NPs (Noun Phrase Ellipsis, *NPE*) that have nominal antecedents with the goal of revealing the mechanisms underlying the retrieval of information associated with the antecedent. In NPE, parts of the nominal phrase are not overtly pronounced. In (2), *key to the cells* is missing from the NP introduced by *Mary's* in the second conjunct, meaning that the interpretation of the missing portion, the ellipsis-site (*NPE-site*) is dependent on an NP in the first conjunct (the *antecedent*), [NP *key to the cells*]. Thus, when an NPE-construction like (2) is processed, the parser needs to "recover" content into the NPE-site by referring to the content of the antecedent.

(2) Derek's key to the cells must be on the table and Mary's [NP ø] is on the carpet.

Anaphoric *one* is another anaphoric construction; like NPE, the interpretation of anaphoric *one* is dependent on an antecedent NP in the first conjunct, ([NP *key to the cells*]), as illustrated in (3) (Hornstein & Lightfoot, 1981; Lidz, Waxman, & Freedman, 2003; Pearl & Lidz, 2013, among others). Thus, it is plausible that when an anaphoric *one* is processed, the parser accesses and retrieves the antecedent of the anaphoric *one*.

(3) Derek's key to the cells must be on the table and Mary's dull one (= key to the cells) is on the carpet.

Recovering the content of NPE and anaphoric *one* should both involve accessing and retrieving the content of the antecedent stored in memory (e.g. Martin & McElree, 2008, 2009, 2011). One important question is what is retrieved when the ellipsis site is processed. One possible processing strategy is to retrieve the head at the first stage and retrieve the local noun (i.e. the modifier in our study) only if necessary. Another possible strategy is to retrieve only the features of the antecedent NP's head. It is also possible to retrieve any parts of the antecedent that match the features of the retrieval cue. To distinguish between these accounts, we used agreement attraction as a diagnostic for retrieval in the processing of NPE and anaphoric *one*, examining whether a local noun contained within the phrase's antecedent elicits attraction.

Against this background, the current study demonstrates that recovering the content of the NPE-site involves retrieving some of the grammatical and structural information associated with the antecedent, such as the syntactic distinction between head and the modifier. We show that the retrieval process is sensitive to a distinction between the head and the modifier within the antecedent NP when the antecedent is retrieved, leading to the same pattern of agreement attraction as observed with fully overt NPs. We compare the processing of NPE to anaphoric *one* which also needs to refer to an antecedent to establish its interpretation, and to non-anaphoric nouns, e.g. *key* vs *necklace*. This demonstrates that the retrieval process involved in ellipsis processing is different from that involved in non-elliptical nominal anaphora constructions. Specifically, we show that the NPE-processing involves more than just accessing and reactivating the antecedent in memory.

2. Background

Under content-addressable retrieval theories, features (e.g. number, gender, case, etc) that match the retrieval cues of the antecedent are retrieved in parallel (Foraker & McElree, 2007; Kush, 2013; Kush & Phillips, 2014; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Martin & McElree, 2008, 2009, 2011; McElree, 2000; McElree, Foraker, & Dyer, 2003; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011).

One piece of evidence for content-addressable memory in parsing is that intervenors that match features of the target item may give rise to processing facilitation, resulting in illusory acceptability of ungrammatical utterances (Dillon, Mishler, Sloggett, & Phillips, 2013; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Parker & Phillips, 2017; Pearlmutter, Garnsey, & Bock, 1999; Tanner, Nicol, & Brehm, 2014; Thornton & MacDonald, 2003; Wagers, Lau, & Phillips, 2009). Previous studies have found that ungrammatical verbs followed by a linearly local but grammatically irrelevant non-head local noun in the modifier in the NP incur less processing costs and improve acceptability ratings for sentences with subject-verb *dis*agreement (Dillon et al., 2013; Lago et al., 2015; Nicol, Forster, & Veres, 1997; Parker & Phillips, 2017; Pearlmutter et al., 1999; Tanner et al., 2014; Tanner, Grey, & van Hell, 2017; Thornton & MacDonald, 2003; Wagers et al., 2009). For example as in (4), the retrieval cue from the verb would trigger the retrieval of a plural subject. Due to the mismatch in number-features the target and the retrieval cue (i.e. the head noun *key* is singular, but the verb *are* is plural), mis-retrieval of the grammatically incompatible element (*boxes*) in the modifier phrase often takes place.

(4) *The key to the boxes are on the table.

This phenomenon is often called *agreement attraction*; under a content-addressable memory framework, it can be viewed as an interference effect where the retrieval of the syntactically illicit elements other than the target results from a partial featural match with the retrieval cues (Dillon et al., 2013; Lago et al., 2015; Nicol et al., 1997; Parker & Phillips, 2017; Tanner et al., 2014; Thornton & MacDonald, 2003). This facilitatory effect exhibits variability based on dependency types: while it has been robustly detected in subject-verb agreement, it has not been as rigorously observed in reflexive processing (Dillon et al., 2013; see also Clifton, Frazier, & Deevy, 1999; Parker & Phillips, 2017; Patson & Husband, 2016; Pearlmutter et al., 1999; Phillips, Wagers, & Lau, 2011; Sturt, 2003; Tanner et al., 2014).

Another point in favour of the role of content-addressable memory in parsing comes from studies of the time-course and accuracy of memory retrieval. Content-addressable retrieval can be characterised by two components: a decrease in comprehension accuracy based on the linear distance between the dependent element and the controlling element, and constant retrieval speed regardless of the complexity of the controlling element. The longer the distance between the dependent element and the controlling element, the lower the comprehension accuracy becomes, due to the increasing number of intervening items (Foraker & McElree, 2007; Martin & McElree, 2008, 2009, 2011; McElree, 2000; McElree et al., 2003; McElree & Dosher, 1989; Van Dyke & McElree, 2011). Because items are

accessed directly, retrieval speed is also predicted to be constant over time regardless of the number of the interpolated items (e.g. words) or the size of search space (e.g. the linear length or structural complexity). These findings are supported by the processing of ellipsis constructions in a Speed-Accuracy Tradeoff (SAT) paradigm (Martin & McElree, 2008, 2009, 2011). Similar results obtain in SAT paradigms for Sluicing (Martin & McElree, 2011) and other dependencies (Foraker & McElree, 2007; McElree, 2000; McElree et al., 2003).

As reviewed above, agreement attraction in comprehension seems to largely occur based on cue-based retrieval mechanisms. However, it is restricted in such a way that erroneous agreement between the verb and non-head noun only occurs in ungrammatical sentences (Lago et al., 2015; Parker & Phillips, 2017; Tanner et al., 2014; Wagers et al., 2009; though, some studies do report that agreement attraction occurs even in grammatical constructions: Acuña-Fariña, Meseguer, & Carreiras, 2014; Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008; Pearlmutter et al., 1999; Parker et al., (2017) and Lago et al. (2015) found an effect in grammatical conditions in some experiments). For example, Wagers et al. (2009) reported that although both (5c) and (5d) are ungrammatical, (5d) is read faster at the verb region and rated more acceptable than (5c) due to the retrieval of the number-matching local noun, with no difference in terms of reading times or acceptability ratings observed in the grammatical (5a) and (5b) (Wagers et al., 2009).

(5) a. The key to the cabinet unsurprisingly was rusty …
   b. The key to the cabinets unsurprisingly was rusty …
   c. *The key to the cabinet unsurprisingly were rusty …
   d. *The key to the cabinets unsurprisingly were rusty …

The implication is that the parser appeals to a cue-based retrieval mechanism to find a controlling element only in reanalysis. The reanalysis process (the process involved in repairing subject-verb disagreement) is instantiated only when the computation of the agreement between the head noun and the verb fails and the parser needs to find a noun that has the same number feature as the verb elsewhere (Lago et al., 2015; Parker & Phillips, 2017; Tanner et al., 2014).

The asymmetric manifestation of agreement attraction suggests the parser's sensitivity to the grammatical distinction between the head and the modifier, i.e. that the parser initially computes number agreement

between the verb and the head noun of the subject, ignoring the local noun. Lago et al. (2015) suggest that this relates to how the structure is predicted by the parser when the subject NP is processed. That is, the number agreement morphology of the verb is predicted when the head noun of the subject NP is identified and processed. If the head noun of the subject is singular, a singular verb is predicted, but if the head noun is plural, a plural verb is predicted. This mismatch can trigger mis-retrieval of a feature-matching local noun.

This asymmetry in attraction based upon predicted and retrieved structure can in turn be used to diagnose what is retrieved when the ellipsis site of the NPE-construction (the NPE-site) is processed. There are at least three possible scenarios with regards to what information associated with the antecedent is retrieved.

2.1. Possibility 1: retrieving the head first and retrieving the local noun when agreement fails.

When the NPE-site is processed, if the head of the antecedent NP is retrieved first and the modifier is retrieved only in cases where the agreement fails, then we expect exactly the same asymmetry of agreement attraction in NPE as observed with other NPs, i.e. attraction effects only in ungrammatical conditions. If agreement attraction is modulated by the grammatical distinction between the head and the modifier, feature-matching local nouns will be accessed and activated only when the number of the verb and the head mismatch and the ungrammatical agreement is recognised.[1]

(6) a. *Derek's key to the box must be on the table and Mary's [NPE key to the box] possibly are on the carpet.
   b. *Derek's key to the boxes must be on the table and Mary's [NPE key to the boxes] possibly are on the carpet.
   c. Derek's key to the box must be on the table and Mary's [NPE key to the box] possibly is on the carpet.
   d. Derek's key to the boxes must be on the table and Mary's [NPE key to the boxes] possibly is on the carpet.

Under this scenario, in (6), once the NPE-site is processed, the parser first retrieves the head noun ([head-N key]) as it is the most prominent element and controls the grammatical and semantic status of the NP. However, if a plural verb is encountered (6a-b), the parser could start looking for another noun that matches the verb number. If the retrieved local noun and the verb match in number, (6b), then the processing of the verb would be facilitated. On the other hand, if

they do not match in number, (6a), then a mismatch cost would be incurred. If the head noun and the local NP are both retrieved when agreement fails, the NPE site should be treated in the same way as an overt NP with the same structure, with a distinction drawn between the head and modifiers. Under this scenario, similar agreement attraction effects are not expected in grammatical conditions, (6c) and (6d) as the agreement is successfully licensed at first pass.

2.2. Possibility 2: retrieving the antecedent without the distinction between the head & the modifier

The second possibility is that different types of features associated with each noun are accessed without a distinction made between the head and the modifier. When processing NPE, all features that overlap with the retrieval cue- whether on the head or modifiers- might be accessed and activated in memory. Items with similar features are likely to be subject to interference effects (so-called similarity-based interference effects; Gordon, Hendrick, & Johnson, 2001, 2004, 2006; Lewis, 1996; Lewis et al., 2006; Lewis & Vasishth, 2005; Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006, 2011), leading to attraction in both grammatical (6d) and ungrammatical (6a) NPE cases.

2.3. Possibility 3: retrieving the head noun only

Third, it is plausible that while processing NPE, the parser retrieves only the information of the head noun of the antecedent NP because the head noun is the locus of the meaning for whole NP and the most prominent element within it (Dillon et al., 2013). If the parser treats the NPE site like an NP that has only the head noun contained within it, no agreement attraction will take place, leading to no acceptability rating or reading time amelioration in 6b/6d vs 6a/6c. However, reading times will be slower in both ungrammatical conditions (6a and 6b) due to the number mismatch.

Here, we have illustrated three different possibilities in terms of what information in the antecedent might be retrieved when the NPE site is processed; each has unique outcomes in terms of acceptability judgements and reading times. This makes searching for agreement attraction in NPE contexts a useful diagnostic for the morphological features of the head and the modifier and the retrieved structure in ellipsis and other anaphoric constructions.

2.4. Overview of the experiments

In order to disclose what is retrieved during the processing of elided and anaphoric elements, six experiments were designed to contrast agreement attraction in NPE with overt NPs (Experiments 1 and 2) and nominal anaphora (Anaphoric *one;* Experiment 3). These included 3 offline acceptability rating experiments (Experiments 1a, 2a, and 3a) and 3 self-paced moving window reading experiments (Experiments 1b, 2b, and 3b).

3. Experiment 1a/1b

These experiments tested whether NPE shows a similar processing profile as non-elliptical NPs. We predict the following: if the antecedent-retrieval process is sensitive to the distinction between the head and the modifier and retrieves the modifier when an ungrammatical verb is detected, facilitation should occur similarly for ungrammatical verbs followed by plural local nouns in NPE and baseline, non-NPE contexts. This would lead to higher acceptability ratings for ungrammatical sentences with plural local nouns (vs singular local nouns) in Experiment 1a and faster reading times for ungrammatical sentences with plural local nouns (vs singular local nouns) in Experiment 1b. However, if the parser uses number features on both the head and the local noun in the modifier as cues to guide antecedent retrieval at the NPE-site, we expect ungrammatical sentences with plural local nouns to be judged more acceptable (Experiment 1a) and read faster (Experiment 1b) in grammatical and in ungrammatical conditions alike. Finally, if only the head is ever retrieved, we expect NPE items to lead to no attraction in acceptability ratings (Experiment 1a) or in reading times (Experiment 1b).

3.1. Experiment 1a NPE: acceptability judgment task (offline)

3.1.1. Participants, materials and design
All 47 participants were native speakers of English with IP addresses from the US and were solicited via Amazon Mechanical Turk (AMT) marketplace. All participants provided informed consent and were compensated $2 for their participation. No participants were excluded.

Critical items consisted of 32 sentence sets arranged in a $2 \times 2 \times 2$ within-subjects factorial design, in which *Local noun number* (singular vs. plural), *Grammaticality* (grammatical vs. ungrammatical), and *NPE* (NPE vs. No NPE) were manipulated as independent factors. All head nouns were singular. A sample set of stimuli is summarised in Table 1. The first five words of each experimental item in the first conjunct always followed the form shown in Table 1 (e.g. *Derek's key to the box/ boxes).* The second conjunct varied by condition. In the baseline conditions, the NP in the first conjunct was

Table 1. Sample stimuli for Experiment 1.

	Factors		
Local Noun	Grammaticality	NPE	Examples
Plural	Grammatical	NPE	… Mary's probably is on the carpet.
Plural	Ungrammatical	NPE	… Mary's probably are on the carpet.
Singular	Grammatical	NPE	… Mary's probably is on the carpet.
Singular	Ungrammatical	NPE	… Mary's probably are on the carpet.
Plural	Grammatical	Baseline	… Mary's key to the boxes probably is on the carpet.
Plural	Ungrammatical	Baseline	… Mary's key to the boxes probably are on the carpet.
Singular	Grammatical	Baseline	… Mary's key to the box probably is on the carpet.
Singular	Ungrammatical	Baseline	… Mary's key to the box probably are on the carpet.

Derek's key to the box/boxes can be on the cabinet and …

repeated (e.g. *Mary's key to the box/boxes*) while in NPE cases, the NP in the first conjunct was elided (e.g. *Mary's*). The first conjunct used a modal verb so as to minimise cues to agreement; the second conjunct included an adverb to isolate effects caused by the local noun from those caused by the verb (see Wagers et al., 2009). The 32 sets of eight conditions were distributed in a pseudo-randomized manner to ensure that participants did not get two experimental items of the same type in a row. The experimental items were combined with 70 grammatical filler sentences of similar length.

3.1.2. Procedure
The IBEX Farm internet-based experimental presentation platform (Drummond, 2011) was used to present the stimuli. For each stimulus, participants observed a single sentence on the screen. Their task was to click on one of the numbered buttons that indicate a 7-point scale, where 1 indicated totally unacceptable and 7 totally acceptable. Ten practice items were presented before presenting the target items.

3.1.3. Analysis
Data were analyzed with linear mixed effect regression using the *lme4* package in R version 3.2.3 (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Bates, Maechler, Bolker, & Walker, 2014; Jaeger, 2008).[2] Each model included simple difference sum-coded fixed effects of *Local noun number* (singular vs. plural), x *Grammaticality* (grammatical vs. ungrammatical), and NPE (whether the sentences involved NPE vs baseline) and their interactions. All models contained the maximal random effects structure justified by the data (Barr, Levy, Scheepers, & Tily, 2013), including random intercepts for participants and items and random slopes for fixed effects where they converged; see model tables for random effect structures. Fixed effects were considered to reach at the significant level at alpha = 0.05 when the absolute value of the t statistic was above 2 (Baayen, 2008).

3.1.4. Results
Mean acceptability scores are shown in Table 2, and mixed effect model outputs are shown in Table 3. All three factors disclosed main effects. A main effect of local noun was observed such that items with ungrammatical singular local nouns were rated lower than their plural counterparts. A main effect of grammaticality was observed such that ungrammatical items were rated significantly less acceptable than grammatical ones. Finally, a main effect of NPE was observed such that items with non-elided NPs were rated significantly less acceptable than those containing NPE.

Effects of grammaticality were qualified by two interactions. An interaction between local noun number and grammaticality was observed such that sentences with singular local nouns were rated less acceptable than sentences with plural local nouns in ungrammatical condition but received equivalent acceptability ratings in

Table 2. Mean acceptability ratings from Experiment 1a.

	Factors		
Local Noun	Grammaticality	Ellipsis	Average raw rating (SE)
Plural	Grammatical	NPE	4.67 (0.16)
Plural	Ungrammatical	NPE	4.28 (0.12)
Singular	Grammatical	NPE	4.67 (0.19)
Singular	Ungrammatical	NPE	3.88 (0.12)
Plural	Grammatical	Baseline	4.41 (0.16)
Plural	Ungrammatical	Baseline	3.55 (0.17)
Singular	Grammatical	Baseline	4.60 (0.13)
Singular	Ungrammatical	Baseline	3.07 (0.16)

Standard errors are in parentheses.

Table 3. Summary of fixed effects from linear mixed effects model in Experiment 1a.

	Estimate	SE	t	p
(Intercept)	4.14	0.16	25.49	
Local Noun	0.17	0.08	2.01	0.05
Grammaticality	−0.90	0.17	−5.39	<0.001***
NPE	0.49	0.13	3.64	<0.001***
Local Noun × Grammaticality	0.52	0.18	2.95	<0.01**
Grammaticality × NPE	0.65	0.19	3.42	<0.01**
Local Noun × NPE	0.06	0.15	0.39	0.70
Local Noun × Grammaticality × NPE	−0.27	0.29	−0.91	0.37

*** p < .001. ** p < .01. * p < .05.
Random intercepts were included for subjects and items, as were by-subject intercepts for Local Noun, Grammaticality, NPE and Local Noun × Grammaticality, and by-item intercepts for Local Noun, Grammaticality, NPE and NPE × Grammaticality.
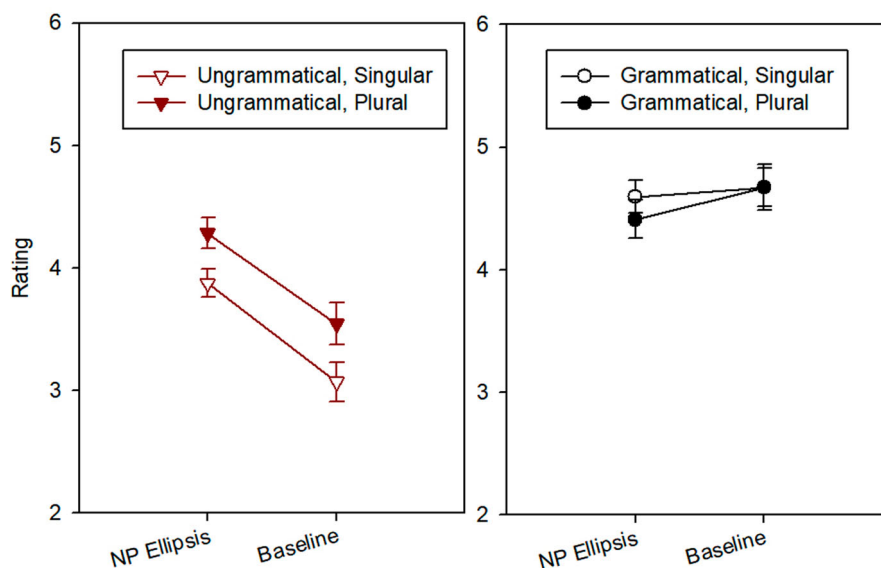
Figure 1. Mean acceptability ratings from Experiment 1a. Error bars indicate standard error.

grammatical conditions. This was confirmed with a subset analysis that revealed a main effect of local noun ($\beta = 0.42$, $SE = 0.12$, $t = 3.61$, $p < 0.001$) in ungrammatical conditions only. An interaction between NPE and grammaticality was also observed such that baseline conditions were judged to be significantly less acceptable than NPE constructions in ungrammatical sentences only. This was confirmed with a subset analysis that revealed a main effect of NPE ($\beta = 0.81$, $SE = 0.16$, $t = 5.12$, $p < 0.001$) in ungrammatical conditions only. Critically, no interactions were observed between local noun and NPE, or between local noun, NPE, and grammaticality. This suggests that the illusion of grammaticality was statistically equivalent whether the NP constituent was overt or elided (Figure 1).

3.2. Experiment 1b NPE: self-paced word-by-word moving window experiment

3.2.1. Participants, materials and design
Participants were 82 native speakers of English from Northwestern University with no history of language disorders. All participants provided informed consent and were either compensated $8/30 minutes or received credit in introductory Linguistics classes; no participants were excluded. Items similar to Experiment 1a were used (see Table 1); some items used final phrases containing other types of constructions (e.g. "*safe in the drawer*") to provide a more varied set of materials to participants. The 32 sets of eight conditions were distributed in a pseudo-randomized manner, and combined with 74 grammatical filler sentences of similar length. The full item sets are available in Appendix A.

3.2.2. Procedure
Stimuli were presented on a desktop PC using Linger software (Rohde, 2003). The experiment followed a self-paced word-by-word moving window paradigm (Just, Carpenter, & Woolley, 1982). Each trial began with dashes masking the words in the sentence. Participants pressed the space bar to display each word as they read. Participants were instructed to read the sentences at a normal speed and to answer the comprehension questions after reading each sentence. The yes/no comprehension question asked participants to press F (yes) or J (no) keys. The critical comprehension questions differed, ranging from "Was the drawer mentioned in the story?" to "Was Anna's brush usually damp after the rain?". The comprehension questions also varied by asking approximately half of the questions related to the first conjunct, and another half to the second conjunct. They were provided with instant feedback about their accuracy. Six practice items were given to participants at the beginning of the experiment so that they became familiarised with the procedure. The experiment took each participant an average of approximate 30 minutes to complete.

3.2.3. Analysis
Following Kazanina, Lau, Lieberman, Yoshida, and Phillips (2007), reading times that exceeded a threshold of 2.5 standard deviations above a participant's mean reading rate for each region were replaced by the threshold value. Dependent measures were identical to Experiment 1a. The regions used for analysis consisted of single words. The critical regions were the verb, the
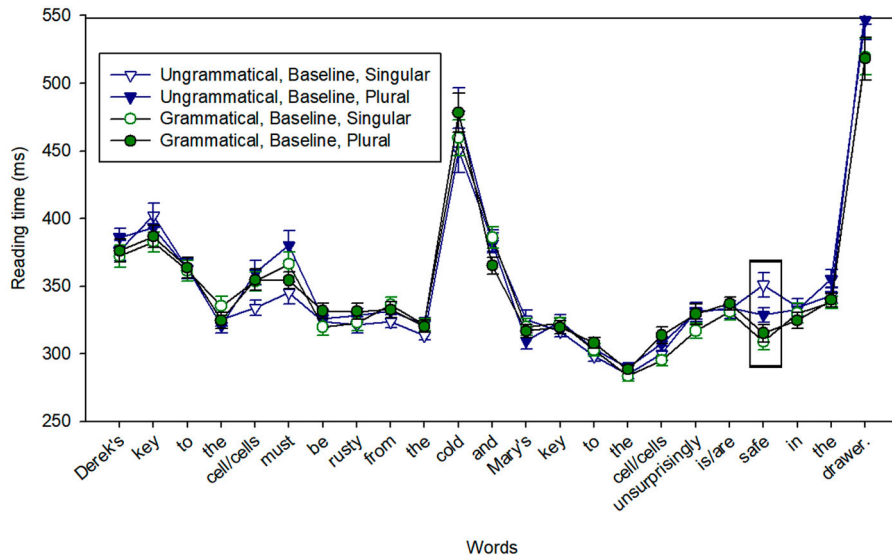
Figure 2. Region-by-region reading time means from Experiment 1b for baseline conditions. Error bars indicate the standard error. The regions of interest are is/are (verb), safe (spillover 1), and in (spillover 2).

following word (spillover region 1) and one word after the spill over region 1 (spillover region 2).

3.2.4. Results

The region-by-region reading times for baseline conditions are presented in Figure 2; those for NPE sentences are presented in Figure 3. Reading times at the critical spillover region for both NPE and baseline conditions are presented in Figure 4 and mixed effect model outputs are presented in Table 4. Mean accuracy for critical trial comprehension questions was 80.0%.

At the verb region, only a main effect of NPE was observed such that items containing NPE were read slower than those without NPE. No other effects were observed.

At the verb spillover region 1, a main effect of grammaticality was observed such that ungrammatical sentences were read slower than grammatical sentences. The critical interaction between local noun and grammaticality was also observed such that the difference between plural local nouns and singular local nouns was larger in ungrammatical sentences than grammatical ones. A subset analysis confirmed that the main effect of local noun was present only in ungrammatical conditions ($\beta = -18.66$, $SE = 6.25$, $t = -2.99$, $p < 0.01$). An interaction between NPE and grammaticality was also observed, such that the difference between NPE and
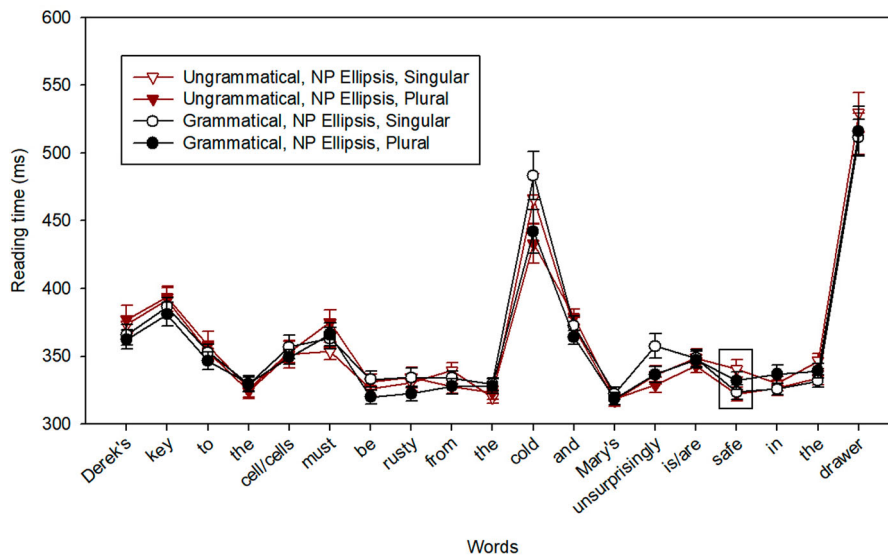

Figure 3. Region-by-region reading time means in Experiment 1b for NPE conditions. Error bars indicate the standard error. The regions of interest are is/are (verb), and safe (spillover 1), and in (spillover 2).
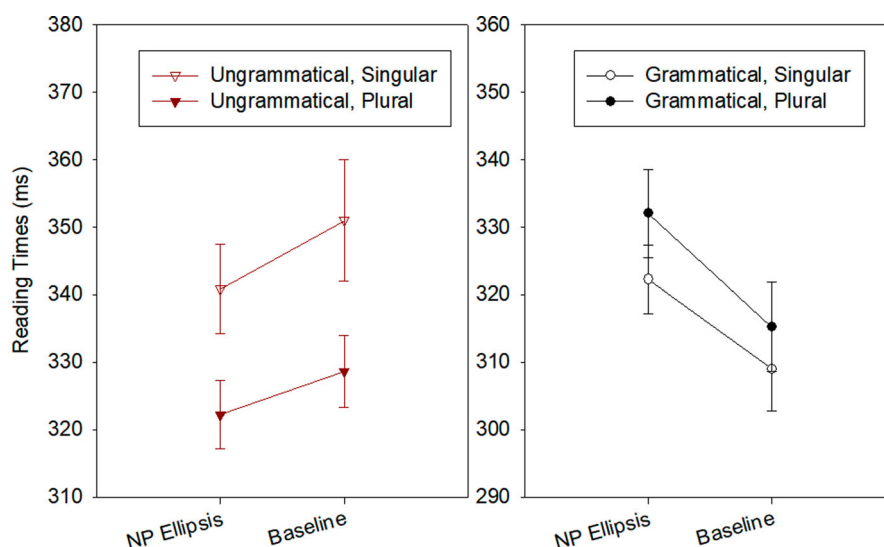
Figure 4. Reading times at the spillover region 1 (safe) for all conditions in Experiment 1b. Error bars indicate the standard error.

Table 4. Summary of results of linear mixed effects models by region in Experiment 1b.

	Estimate	SE	t	p
Verb Region (*is/are*): by-subject random intercepts and slopes for Local Noun, NPE and Grammaticality, by-item random intercepts and slopes for NPE and Grammaticality				
(Intercept)	338.96	8.79	38.56	
Local Noun	1.17	4.02	0.29	0.77
Grammaticality	−0.90	4.54	−0.20	0.84
NPE	12.58	4.65	2.71	<0.05*
Local Noun * Grammaticality	−5.18	7.73	−0.67	0.50
Grammaticality * NPE	0.17	7.73	0.02	0.98
Local Noun * NPE	−7.02	7.73	−0.91	0.36
Local Noun * Grammaticality * NPE	1.58	15.46	0.10	0.92
Verb Spill-over Region 1 (*safe*): by-subject random intercepts and slopes for Grammaticality and NPE, by-item random intercepts and slopes for NPE				
(Intercept)	326.47	8.42	38.78	
Local Noun	−5.56	3.66	−1.52	0.13
Grammaticality	15.09	5.15	2.93	<0.01**
NPE	3.17	4.46	0.71	0.48
Local Noun * Grammaticality	−25.88	7.33	−3.53	<0.001***
Grammaticality * NPE	−22.37	7.32	−3.05	<0.01**
Local Noun * NPE	1.78	7.32	0.24	0.81
Local Noun * Grammaticality * NPE	0.91	14.65	0.06	0.95
Verb Spill-over Region 2 (*in*): by-subject random intercepts and slopes for Local Noun, NPE and Grammaticality, by-item random intercepts and slopes for NPE and Grammaticality				
(Intercept)	328.69	7.49	47.91	
Local Noun	0.98	3.72	0.26	0.79
Grammaticality	2.40	3.90	0.62	0.54
NPE	−1.84	4.31	−0.43	0.67
Local Noun * Grammaticality	−2.45	7.13	−0.34	0.73
Grammaticality * NPE	−9.18	7.13	−1.29	0.20
Local Noun * NPE	2.65	7.13	0.37	0.71
Local Noun * Grammaticality * NPE	−15.24	14.26	−1.07	0.29

*** $p < .001$. ** $p < .01$. * $p < .05$.

the baseline was larger for grammatical sentences. Critically, there was no main effect of NPE, and no interaction between any other factors, suggesting that items containing plural local nouns were always facilitated at the verb spillover region, regardless of NPE. This further suggests that NPE and the no-ellipsis baseline were treated similarly under conditions that elicit attraction. However, it is possible that the absence of a three-way interaction might be due to insufficient statistical power and that such a possibly small effect could be uncovered if even more data were collected. Finally, at the verb spillover region 2, no effects were significant; all conditions were processed similarly.

3.2.5. Discussion

Experiment 1a and 1b aimed to address the nature of attraction effects in NPE in offline and online tasks. Experiment 1a showed that sentences with ungrammatical plural local nouns were judged more acceptable than those with ungrammatical singular local nouns regardless of NPE, with no significant difference in acceptability ratings in grammatical conditions. Experiment 1b revealed attraction effects in NPE and in baseline constructions following ungrammatical verbs, with no corresponding evidence of attraction in similar grammatical sentences.

These results are most compatible with an account where the head noun is initially retrieved at the NPE-site and the local noun is retrieved when triggered by ungrammatical agreement. This means that the parser distinguishes the head and the local noun in the elided phrase. This supports the view that the grammatical information associated with an antecedent is retrieved within the NPE site. With NPE as with overt NPs, the parser uses this information in a reanalysis process with a cue-based retrieval mechanism only after the apparent detection of a mismatch in number agreement (Lago et al., 2015; Parker & Phillips, 2017; Tanner et al., 2014; Wagers et al., 2009).

Reading time effects of verb ungrammaticality were reflected relatively late in processing, appearing at the spillover region 1 and not at the critical verb region. This suggests that the antecedent search and retrieval for NPE is influenced by the availability of cues. In order to recognise the NPE-site, the parser needs to first recognise that *Mary's* and *probably* are incompatible and needs to insert a silent NP between them, triggering the need to do antecedent retrieval before the verb has been processed. The lack of a role for morphology in guiding antecedent retrieval may be a result of the fact that in NPE, the elided NP is silent and thus does not provide overt morphological cues. The reanalysis process for the parser to recognise the NPE-site may contribute to the processing complexity. This may mask the grammaticality effect at the verb region. We return to the fine-grained time profiles of the error detection and the reanalysis processes involved in resolving ellipsis and non-elliptical nominal anaphora constructions in the Discussion session.

Finally, our results are not compatible with the hypothesis that only the head noun is retrieved at the ellipsis site[3]. If only the head noun were retrieved, we would expect to observe similarity-based interference in grammatical cases, slowing singular local nouns relative to plurals. However, our results revealed no differences between plural and singular local nouns in grammatical conditions. The results here are also incompatible with the hypothesis that the parser retrieves the content of the antecedent without recourse to the grammatical properties of the antecedent. Under this scenario, no distinctions between the head and the modifier ($[_{PP}$ to $[_{DP}$ the $[_{NP}$ box/boxes]]]]) are drawn when the antecedent is accessed. According to this hypothesis, we would thus expect agreement attraction in grammatical and ungrammatical cases, as features are retrieved in parallel.

Taken together, we conclude that agreement attraction in NPE is most consistent with the scenario in which the head is retrieved initially and the local noun is retrieved only when the head noun and the verb do not agree[4]. Other hypotheses appealing to retrieval of only the head or the content without the distinction between the head and the modifier fail to explain why we observe agreement attraction in ungrammatical conditions regardless of NPE.

However, there is an alternate explanation of the observed data which attributes the NPE effects to the nature of the coordination structure itself. A growing body of research suggests that the parallel structure in *and*-coordinated sentences facilitates the access and reactivation of the elements in the first conjunct which are maintained as active in memory until the elements

in the second conjunct are encountered (Callahan, Shapiro, & Love, 2010; Arregui, Clifton, Frazier, & Moulton, 2006; Dickey & Bunger, 2011; Frazier, Munn, & Clifton, 2000; Kehler, 2000; Poirier, Wolfinger, Spellman, & Shapiro, 2010; Shapiro, Hestvik, Lesan, & Garcia, 2003; Sturt, Keller, & Dubey, 2010; Tanenhaus & Carlson, 1990). Thus, one may argue that coordination with parallel conjuncts is sufficient to elicit a search for a matching feature in the first conjunct, resulting in agreement attraction. Experiment 2 was designed to test this possibility

4. Experiment 2a/2b

The goal of Experiment 2 is to investigate whether the agreement attraction observed in NPE contexts is due to ellipsis or to the coordinated context itself. This was tested by replacing the anaphoric element with an entirely different noun as in (7).

(7) Derek's **key to the cell/s** must be rusty from the cold, and Mary's **necklace/necklaces** probably is/are safe in the drawer.

We predict the following: if ellipsis or another anaphoric element is crucial for the effects observed in Experiment 1, then we predict no agreement attraction effect in the No Anaphora conditions, because there are no elided nouns or anaphoric elements in the second conjunct and there is no dependency waiting to be resolved. This would lead to minimal differences in acceptability ratings for ungrammatical sentences with plural versus singular local nouns in the No Anaphora condition in Experiment 2a and to similar reading times in ungrammatical sentences with plural versus singular local nouns in the No Anaphora condition in Experiment 2b. However, if coordination is sufficient to trigger an agreement attraction effect regardless of anaphora, we expect the No Anaphora ungrammatical sentences with plural local nouns *(key to the cells … necklaces is)* to be judged more acceptable in Experiment 2a and read faster in Experiment 2b than their singular counterparts *(key to the cell … necklace are)*.

4.1. Experiment 2a No Anaphora: acceptability judgment task (offline)

4.1.1. Participants, materials and design
Participants were 60 native speakers of English from the Northwestern University community with no history of language disorders. All participants provided informed consent and were compensated $8/30 minutes or received credit in an introductory Linguistics class. No participants were excluded.

Table 5. Sample stimuli for Experiment 2.

Factors			
Local Noun	Grammaticality	No Anaphora	Examples
Plural	Grammatical	No Anaphora	… Mary's necklaces probably are safe in the drawer.
Plural	Ungrammatical	No Anaphora	… Mary's necklaces probably is safe in the drawer.
Singular	Grammatical	No Anaphora	… Mary's necklace probably is safe in the drawer.
Singular	Ungrammatical	No Anaphora	… Mary's necklace probably are safe in the drawer.
Plural	Grammatical	Baseline	… Mary's key to the cells probably is safe in the drawer.
Plural	Ungrammatical	Baseline	… Mary's key to the cells probably are safe in the drawer.
Singular	Grammatical	Baseline	… Mary's key to the cell probably is safe in the drawer.
Singular	Ungrammatical	Baseline	… Mary's key to the cell probably are safe in the drawer.

Derek's key to the cell/cells must be rusty from the cold and …

Critical items consisted of 32 sentence sets arranged in a $2 \times 2 \times 2$ within-subjects factorial design, in which *Local noun number* (singular vs. plural), *Grammaticality* (grammatical vs. ungrammatical), and *No Anaphora* (No Anaphora vs. Baseline) were manipulated as independent factors. Experimental items were similar to those used in Experiment 1 except that instead of NPE, the noun in the baseline condition was substituted with an alternate noun in order to eliminate the anaphoric element in the first conjunct. A sample set of stimuli is summarised in Table 5. Items were distributed in a pseudo-randomized manner to ensure that participants did not get two experimental items of the same type in a row. The experimental items were combined with 50 grammatical filler sentences of similar length.

4.1.2. Procedure
Stimuli were presented on a desktop PC using Linger software (Rohde, 2003). The task was otherwise identical to Experiment 1a.

4.1.3. Analysis
The analysis was similar to Experiment 1a[5]. Each model included simple difference sum-coded fixed effects of local noun (whether the local noun was plural or singular), grammaticality (whether the local noun and the verb matched in number agreement), No Anaphora (whether the sentences involved a new noun with no anaphora vs baseline) and their interactions, as well as random intercepts for participants and items and the maximum number of random slopes justified by the data (Barr et al., 2013).

4.1.4. Results
Mean acceptability scores are shown in Table 6, and a summary of results is shown in Table 7. A main effect of local noun was observed such that items with ungrammatical singular local nouns were rated lower than their plural counterparts. A main effect of No Anaphora was also observed such that items with a new non-anaphoric noun were rated lower than the baseline items.

Table 6. Mean acceptability ratings from Experiment 2a.

Factors			
Local Noun	Grammaticality	No Anaphora	Average raw rating (SE)
Plural	Grammatical	No Anaphora	5.18 (0.14)
Plural	Ungrammatical	No Anaphora	3.07 (0.13)
Singular	Grammatical	No Anaphora	5.13 (0.13)
Singular	Ungrammatical	No Anaphora	2.79 (0.13)
Plural	Grammatical	Baseline	5.28 (0.12)
Plural	Ungrammatical	Baseline	3.51 (0.14)
Singular	Grammatical	Baseline	5.36 (0.13)
Singular	Ungrammatical	Baseline	2.95 (0.13)

Standard errors are in parentheses.

Table 7. Summary of results of linear mixed effects model in Experiment 2a.

	Estimate	SE	t	p
(Intercept)	4.16	0.12	33.96	
Local Noun	0.20	0.06	3.15	< 0.01**
Grammaticality	−2.16	0.21	−10.46	< 0.001***
No Anaphora	−0.23	0.08	−3.04	<0.01**
Local Noun × Grammaticality	0.43	0.14	3.17	<0.001***
Grammaticality × No Anaphora	−0.13	0.11	−1.17	0.24
Local Noun × No Anaphora	−0.07	0.11	−0.62	0.53
Local Noun × Grammaticality × No Anaphora	−0.40	0.22	−1.84	0.07

*** $p < .001$. ** $p < .01$. * $p < .05$.
Random intercepts were included for subjects and items, as were by-subject random slopes for Local Noun, Grammaticality, No Anaphora and Local Noun × Grammaticality, and by-item random slopes for Local Noun, Grammaticality, No Anaphora, and Local Noun × Grammaticality.

Additionally, a main effect of grammaticality was observed such that ungrammatical items were rated significantly lower than their grammatical counterparts. These were qualified by an interaction between local noun and grammaticality such that sentences with singular local nouns were judged less acceptable than those with plural local nouns in ungrammatical conditions, as well as by a marginal three-way interaction such that sentences with plural local nouns were judged to be most acceptable in the ungrammatical baseline condition. No other significant main effects or interactions were observed.

The heightened effects of local noun and No Anaphora in in ungrammatical conditions were supported by a subset analysis. In ungrammatical items, there were main effects of local noun ($\beta = 0.41$, $SE = 0.10$, $t = 3.97$, $p < 0.001$) and No Anaphora ($\beta = -0.30$, $SE = 0.11$,
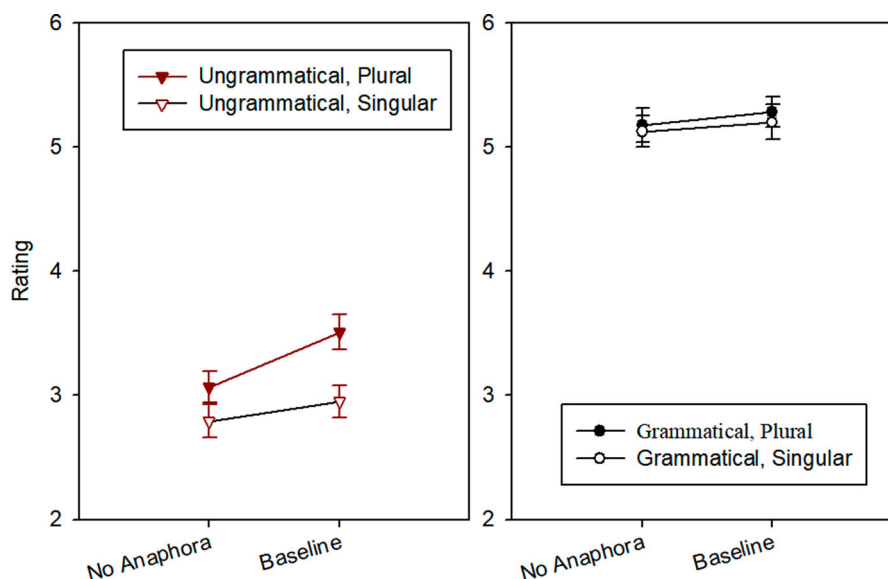
Figure 5. Mean acceptability ratings from Experiment 2a. Error bars indicate standard error.

$t = -2.75$, $p < 0.01$). This confirms that in ungrammatical conditions, singular local nouns and new non-anaphoric nouns led to lower acceptability ratings. In contrast, in grammatical items, only a marginal main effect of No Anaphora was observed ($\beta = -0.17$, $SE = 0.09$, $t = -1.86$, $p = 0.07$) such that items containing new non-anaphoric nouns were judged marginally less acceptable (Figure 5).

4.2. Experiment 2b No Anaphora: self-paced word-by-word moving window experiment

4.2.1. Participants & materials and design
Participants were 78 native speakers of English from the Northwestern University community with no history of language disorders. All participants provided informed consent and received credit in an introductory Linguistics class; no participants were excluded. The same 32 critical items were used as in Experiment 2a; items in the eight conditions were distributed in a pseudo-randomized order and combined with 74 grammatical filler sentences of similar length.

4.2.2. Procedure
The procedure was identical to Experiment 1b.

4.2.3. Analysis
Dependent measures were identical to Experiment 2a and the analysis procedure matched Experiment 1b.

4.2.4. Results
Region-by-region reading times for baseline conditions are presented in Figure 6; those for No Anaphora constructions are presented in Figure 7. Reading times at the critical spillover region for both are presented in Figure 8. Mixed effect model outputs are presented in Table 8. Mean accuracy for critical trial comprehension questions was 80.0%.

At the critical verb region, a main effect of grammaticality was observed such that ungrammatical items were read slower than grammatical ones. A main effect of No Anaphora was also observed such that items with new non-anaphoric nouns were read significantly slower than baseline items.

At the verb spillover region 1, a main effect of grammaticality was observed again, such that ungrammatical items were read slower than grammatical ones. Again, a main effect of No Anaphora was also observed, such that items with new non-anaphoric nouns were read slower than the baseline. These were qualified by an interaction between grammaticality and No Anaphora, such that the difference between the No Anaphora and baseline conditions was larger for ungrammatical verbs, as confirmed by subset analyses (for ungrammatical sentences: $\beta = 29.06$, $SE = 9.82$, $t = 2.96$, $p < 0.01$; for grammatical sentences: ($\beta = 8.75$, $SE = 4.84$, $t = 1.81$, $p = 0.07$).

An interaction between the local noun and No Anaphora was also observed such that local noun number affected the No Anaphora and baseline items differently: while singular local nouns were read more slowly in the baseline conditions, they were read more quickly in the No Anaphora conditions. Splitting on grammaticality shows that local noun by No Anaphora effects were restricted to grammatical items. Grammatical items with new non-anaphoric plural nouns were read most slowly, showing a marginal interaction between No Anaphora and Local Noun ($\beta = 16.09$, $SE = 9.13$, $t = 1.76$, $p = $
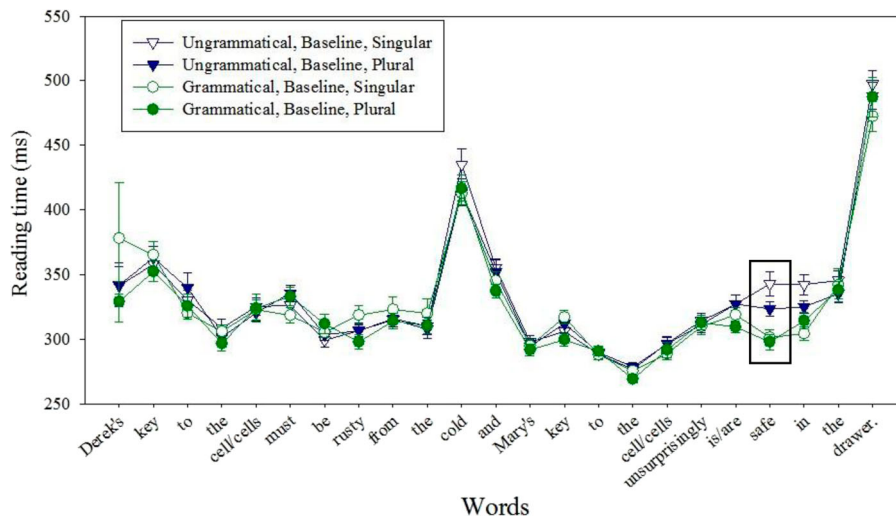
Figure 6. Region-by-region reading time means from Experiment 2b for baseline conditions. Error bars indicate the standard error. The regions of interest are is/are (verb), safe (spillover 1), and in (spillover 2).
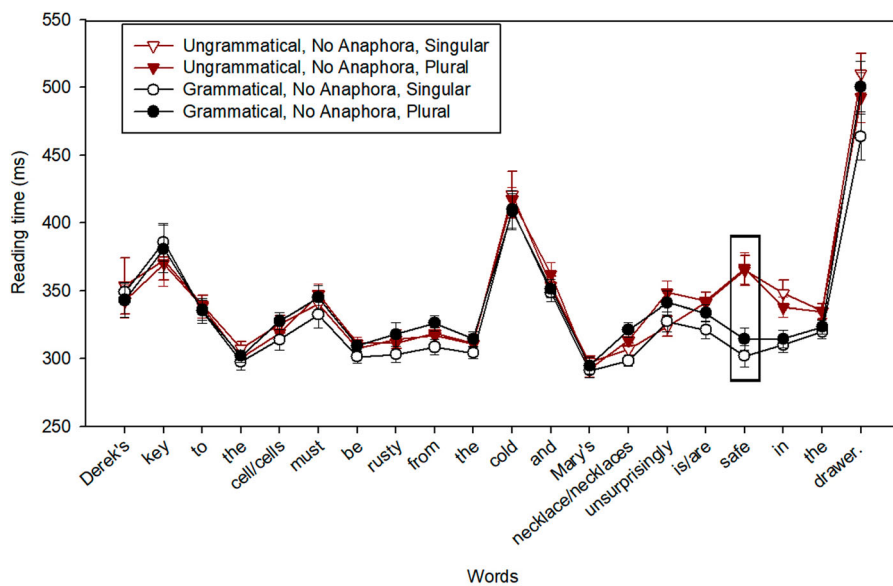

Figure 7. Region-by-region reading time means from Experiment 2b for No Anaphora conditions. Error bars indicate the standard error. The regions of interest are is/are (verb), safe (spillover 1), and in (spillover 2).

0.08) and a marginal main effect of No Anaphora ($\beta =$ 8.75, $SE = 4.85$, $t = 1.80$, $p = 0.07$). In contrast, in ungrammatical items, only a main effect of No Anaphora was observed ($\beta = 29.14$, $SE = 9.92$, $t = 2.94$, $p < 0.01$). This supports the view that while non-anaphoric nouns increased reading times, reading time differences between sentences with plural local nouns and singular local nouns in ungrammatical sentences were minimal.

At the verb spillover region 2, effects of grammaticality were observed in the form of a main effect such that ungrammatical items were read slower than grammatical ones. An interaction between local noun and grammaticality was observed such that items containing local singular nouns and ungrammatical verbs were read

especially slowly; a subset analysis showed that this interaction was largely driven by the baseline conditions as there was a main effect of grammaticality ($\beta = 2\,4\,.1\,4$, $SE = 7.56$, $t = 3.19$, $p < 0.01$) and an interaction between grammaticality and local noun number in the baseline condition ($\beta = -23.80$, $SE = 10.77$, $t = -2.21$, $p < 0.05$) but only a main effect of grammaticality and in the No Anaphora condition ($\beta = 2\,8.17$, $SE = 7.71$, $t = 3.65$, $p < 0.001$).

4.2.5. Discussion (Experiment 2a & 2b)

The goal of Experiments 2a and 2b was to rule out the possibility that agreement attraction in NPE is due to coordination alone. This was done by replacing the
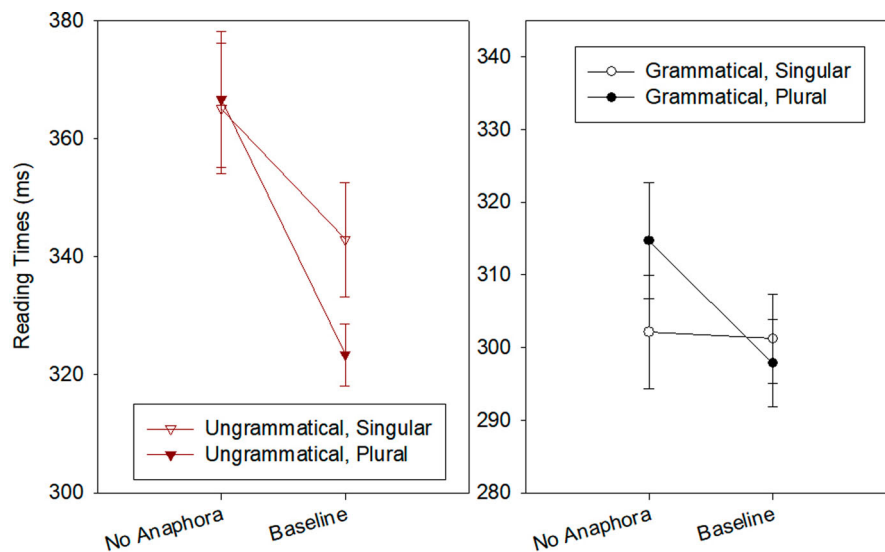
Figure 8. Reading times at the spillover region 1 (safe) for all conditions in Experiment 2b. Error bars indicate the standard error.

Table 8. Summary of results of linear mixed effects models by region in Experiment 2b.

	Estimate	SE	t	p
Verb Region (*is/are*)				
by-subject random intercepts and slopes for Local Noun, Grammaticality, No Anaphora, and by-item random intercepts and slopes for Local Noun, Grammaticality, and No Anaphora				
(Intercept)	325.91	9.03	36.10	
Local Noun	−0.39	4.44	−0.09	0.93
Grammaticality	12.61	5.66	2.23	<0.05*
No Anaphora	12.86	4.68	2.75	<0.01**
Local Noun × Grammaticality	−3.45	8.13	−0.42	0.67
Grammaticality × No Anaphora	1.27	8.13	0.16	0.88
Local Noun × No Anaphora	11.06	8.13	1.36	0.17
Local Noun × Grammaticality × No Anaphora	−16.72	16.27	−1.03	0.30
Verb Spill-over Region 1 (*safe*)				
by-subject random intercepts and slopes for Grammaticality and No Anaphora, and by-item random intercepts and slopes for Local Noun, Grammaticality and No Anaphora				
(Intercept)	325.20	9.46	34.40	
Local Noun	−2.17	5.62	−0.39	0.70
Grammaticality	44.05	8.32	5.30	<0.001***
No Anaphora	19.08	5.77	3.31	<0.01**
Local Noun × Grammaticality	−11.69	9.72	−1.20	0.23
Grammaticality × No Anaphora	20.64	9.72	2.12	<0.05*
Local Noun × No Anaphora	19.41	9.72	2.00	<0.05*
Local Noun × Grammaticality × No Anaphora	6.75	19.45	0.35	0.73
Verb Spill-over Region 2 (*in*)				
by-subject random intercepts and slopes for Local Noun, Grammaticality and No Anaphora, and by-item random intercepts and slopes for Local Noun, Grammaticality, and No Anaphora.				
(Intercept)	323.26	8.79	39.76	
Local Noun	−2.94	5.50	−0.53	0.59
Grammaticality	26.15	6.77	3.86	<0.001***
No Anaphora	6.00	4.25	1.41	0.16
Local Noun × Grammaticality	−18.84	7.70	−2.45	<0.05*
Grammaticality × No Anaphora	4.02	7.70	0.52	0.60
Local Noun × No Anaphora	0.07	7.70	0.01	0.99
Local Noun × Grammaticality × No Anaphora	9.69	15.40	0.63	0.53

*** $p < .001$. ** $p < .01$. * $p < .05$.

anaphoric element with an entirely different noun. If an NP somewhere in the sentence matches the number feature of the verb, and the plural source deriving from the coordinated *and* is strong enough to trigger agreement attraction, higher acceptability ratings and attenuated reading times in No Anaphora conditions would be expected even with no anaphoric element in the second conjunct.

Results of Experiment 2a show that ungrammatical sentences with plural local nouns were rated more

acceptable than ungrammatical sentences with singular local nouns in the baseline condition, replicating Experiment 1a. No significant difference was observed in acceptability ratings within grammatical conditions, nor were significant differences observed between local singular and local plural nouns in the ungrammatical No Anaphora conditions.

Results of Experiment 2b showed agreement attraction in the ungrammatical baseline conditions such that ungrammatical verbs following plural local nouns were read faster than ungrammatical verbs following singular local nouns. This pattern is consistent with the previous study and with the hypothesis that attraction occurs as a result of a reanalysis process in order to reconcile the feature violation between the head noun and the predicted number of the verb (Lago et al., 2015; Parker & Phillips, 2017; Tanner et al., 2014; Wagers et al., 2009).

In contrast, in the No Anaphora condition in Experiment 2b, where the NP in the second conjunct was completely novel, plural and singular local nouns were read similarly quickly at the spillover region. This means that in the No Anaphora condition, even when the readers detect an agreement error (*necklace are, necklaces is*), they do not search for an antecedent in the first conjunct due to the absence of an anaphoric element. In combination with the results of Experiment 1b, this suggests that coordination is not sufficient to trigger agreement attraction, and that either an anaphoric element or ellipsis is required to prompt the retrieval of an antecedent. Although a large body of research suggests that the parallel structure in the coordination context affects the reactivation of the elements in the first conjunct (Arregui et al., 2006; Callahan et al., 2010; Dickey & Bunger, 2011; Frazier et al., 2000; Frazier & Clifton, 2001; Kehler, 2000; Poirier et al., 2010; Shapiro et al., 2003; Sturt et al., 2010; Tanenhaus & Carlson, 1990), we observed that coordination with a parallel conjunct is not sufficient for the parser to look for a feature matching noun in the left-context.

The lack of agreement attraction in the No Anaphora condition in Experiment 2b contrasts with the offline judgment task presented in Experiment 2a, where local plural nouns tended to elicit slightly higher ratings in the ungrammatical No Anaphora condition. The discrepancy between the results from offline and online experiments for the No Anaphora conditions might be attributed to what is available to the parser. In offline judgment tasks, participants are able to rigorously examine the first conjunct to interpret the sentence. Because of this left context readers may have therefore been more susceptible to the interference effect caused by the morphological overlap with the noun in the first conjunct in the offline judgment task.

5. Experiment 3a/3b

The goal of Experiment 3 was to rule out a final alternate account of the data, testing whether the agreement attraction observed in NPE contexts in Experiment 1 was truly due to the retrieval of the antecedent. An alternate possibility is that the parser is simply referring to the antecedent in the first conjunct without actually retrieving any grammatical information at the ellipsis site. To rule this out, Experiment 3 tests whether the antecedent retrieval is grammatically constrained by using an anaphoric element with a strong morphological cue. If the parser is merely accessing the antecedent without making a distinction between the head and the modifier in the NPE-site, the same pattern of agreement attraction is predicted for anaphoric *one* as was observed in Experiment 1. However, if the parser is accessing the antecedent differently for anaphoric *one* compared to NPE, we expect to see no agreement attraction pattern for the following reasons.

As we discussed earlier, like NPE, the interpretation of anaphoric *one* (Crain, 1994; Hornstein & Lightfoot, 1981; Lidz et al., 2003; Lightfoot, 1989; Payne, Pullum, Scholz, & Berlage, 2013) is dependent on an antecedent NP in the first conjunct. Anaphoric *one*, however, differs from NPE in that it provides a strong morphological cue that it refers to a singular NP and the head noun: anaphoric *one* triggers a search for the antecedent, privileging the head noun over the local noun. Previous research as also shown that cue reliability has been proved to be a strong factor in that reliable marking blocks agreement attraction (Franck et al., 2008; Hartsuiker, Schriefers, Bock, & Kikstra, 2003; Vigliocco & Franck, 1999; Vigliocco, Butterworth, & Semenza, 1995). Thus, when anaphoric *one* is processed, it sets up a strong prediction for a singular verb. The prediction is that if anaphoric *one* is processed differently from NPE, it may engender a local ungrammaticality rather than the attraction effect that was observed in Experiment 1. Similarly, differences between anaphoric *one* and baseline construction are also predicted for grammatical conditions, as in the grammatical Anaphoric *one* condition, the parser may easily disregard information on the local noun.

5.1. Experiment 3a anaphoric one: acceptability judgement task (offline)

5.1.1. Participants & materials and design
Participants were 60 native speakers of English from the Northwestern University community with no history of

Table 9. Sample stimuli for Experiment 3.

	Factors		
Local Noun	Grammaticality	Anaphoric one	Examples
Plural	Grammatical	AO	… Mary's dull one unsurprisingly is safe in the drawer.
Plural	Ungrammatical	AO	… Mary's dull one unsurprisingly are safe in the drawer.
Singular	Grammatical	AO	… Mary's dull one unsurprisingly is safe in the drawer.
Singular	Ungrammatical	AO	… Mary's dull one unsurprisingly are safe in the drawer.
Plural	Grammatical	Baseline	… Mary's dull key to the boxes unsurprisingly is safe in the drawer.
Plural	Ungrammatical	Baseline	… Mary's dull key to the boxes unsurprisingly are safe in the drawer.
Singular	Grammatical	Baseline	… Mary's dull key to the box unsurprisingly is safe in the drawer.
Singular	Ungrammatical	Baseline	… Mary's dull key to the box unsurprisingly are safe in the drawer.

Derek's key to the cell/cells must be rusty from the cold and …

language disorders; no participants were excluded. Critical items consisted of 32 sentence sets arranged in a 2 × 2 × 2 within-subjects factorial design, in which *Local noun number* (singular vs. plural), *Grammaticality* (grammatical vs. ungrammatical) and *Anaphoric one* (Anaphoric *one* vs. baseline) were manipulated as independent factors. All head nouns were singular. A sample set of stimuli is summarised in Table 9. Items were similar to Experiments 1 and 2, but contained items with anaphoric *one* (*Mary's one*) rather than NPE or the No Anaphora condition. The 32 sets of eight conditions were distributed in a pseudo-randomized manner, to ensure that participants did not get two experimental items of the same type in a row. The experimental items were combined with 74 grammatical filler sentences of similar length.

5.1.2. Procedure
The procedure was the same as Experiment 2a.

5.1.3. Analysis
Analysis was similar to Experiment 1a and 2a; fixed effects were *Local noun number* (singular vs. plural), *Grammaticality* (grammatical vs. ungrammatical), and Anaphoric *one* (whether the sentences involved AO vs baseline) and their interactions[6].

5.1.4. Results
Mean acceptability scores are shown in Table 10, and mixed effect model outputs are shown in Table 11. A marginal main effect of local noun was observed such

Table 10. Mean acceptability ratings from Experiment 3a.

	Factors		
Local Noun	Grammaticality	Anaphoric One	Average raw rating (SE)
Plural	Grammatical	AO	5.36 (0.12)
Plural	Ungrammatical	AO	3.91 (0.12)
Singular	Grammatical	AO	5.38 (0.13)
Singular	Ungrammatical	AO	3.71 (0.12)
Plural	Grammatical	Baseline	5.40 (0.12)
Plural	Ungrammatical	Baseline	4.30 (0.12)
Singular	Grammatical	Baseline	5.61 (0.11)
Singular	Ungrammatical	Baseline	3.86 (0.13)

Standard errors are in parentheses.

Table 11. Summary of results of linear mixed effects models in Experiment 3a.

	Estimate	SE	t	p
(Intercept)	4.69	0.13	37.12	
Local Noun	0.09	0.05	1.68	0.09
Grammaticality	−1.48	0.18	−8.44	< 0.001***
Anaphoric *one*	−0.19	0.07	−2.89	< 0.05*
Local Noun × Grammaticality	0.45	0.11	4.19	< 0.001***
Grammaticality × AO	−0.10	0.11	−0.92	0.36
Local Noun × AO	−0.05	0.11	−0.48	0.63
Local Noun × Grammaticality × AO	−0.51	0.21	−2.40	<0.05*

*** $p < .001$. ** $p < .01$. * $p < .05$.
Random intercepts were included for subjects and items, as were by-subject random slopes for Grammaticality, and Anaphoric *one*, and by-item random slopes for Grammaticality and Anaphoric *one*.

that items containing singular local nouns were rated less acceptable than those with plural local nouns. A main effect of grammaticality was also observed such that ungrammatical items were rated less acceptable than grammatical ones. These effects were qualified by an interaction between local noun and grammaticality such that ungrammatical items with singular local nouns were rated least acceptable. A three-way interaction between grammaticality, local noun, and Anaphoric *one* was also observed suggesting that items with local singular nouns were rated significantly worse only in the ungrammatical baseline condition. Interactions with grammaticality were confirmed with a subset analysis which showed main effects of local noun ($\beta = 0.30$, $SE = 0.89$, $t = 3.39$, $p < 0.01$), Anaphoric *one* ($\beta = −0.25$, $SE = 0.10$, $t = −2.49$, $p < 0.05$) and an interaction between the two ($\beta = −0.30$, $SE = 0.15$, $t = −2.01$, $p < 0.05$) in ungrammatical conditions; all grammatical conditions received equivalent acceptability ratings (Figure 9).

5.2. Experiment 3b anaphoric one: self-paced word-by-word moving window experiment

5.2.1. Participants & materials and design
Participants were 91 native speakers of English from the Northwestern University community with no history of language disorders. All participants provided informed consent and received credit in an introductory Linguistics class. One participant was excluded because the
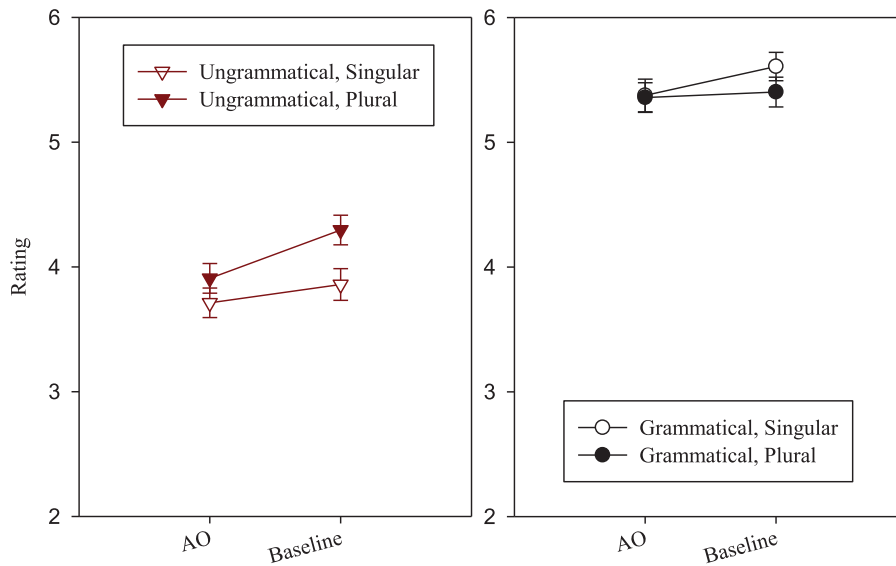
Figure 9. Mean acceptability ratings from Experiment 3a. Error bars indicate standard error.

participant's comprehension question accuracy rate was close to 50%, not significantly better than if they had selected their answer at random. Similar critical items were used as in Experiment 3a (see Table 11), but an adjective was included in the first conjunct to increase the diversity and naturalness of the items. The 32 sets of eight conditions were distributed in a pseudo-randomized manner to ensure that participants did not get two experimental items of the same type in a row. Critical items were combined with 74 grammatical filler sentences of similar length.

5.2.2. Procedure
The procedure was identical to Experiment 1b and 2b.

5.2.3. Analysis
Dependent measures were the same as Experiment 3a, and the analysis procedure matched Experiment 1b and Experiment 2b.

5.2.4. Results
The region-by-region reading times for baseline conditions are presented in Figure 10; those for Anaphoric *one* (AO) constructions are presented in Figure 11. Reading times at the critical spillover region for both are presented in Figure 12. Mixed effect model outputs are presented in Table 12. Mean accuracy for critical trial comprehension questions was 75%.
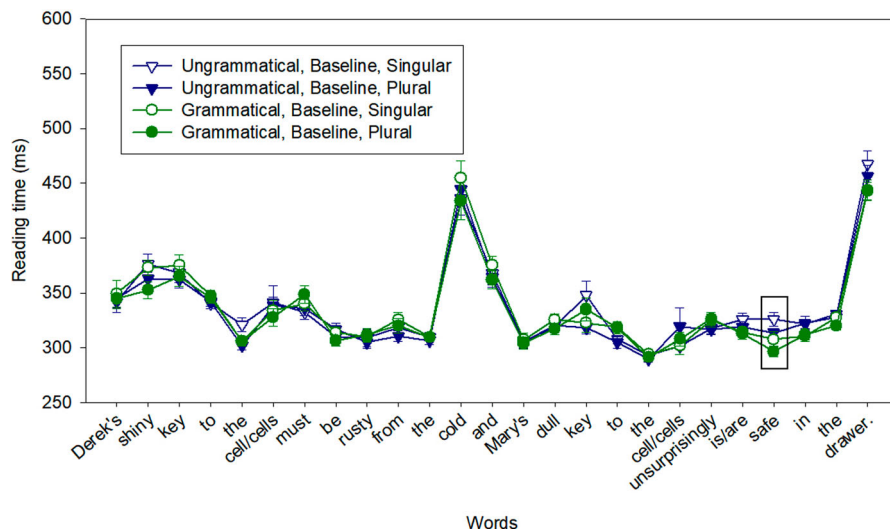

Figure 10. Region-by-region reading time means from Experiment 3b for baseline conditions. Error bars indicate the standard error. The regions of interest are is/are (verb), safe (spillover 1), and in (spillover 2).
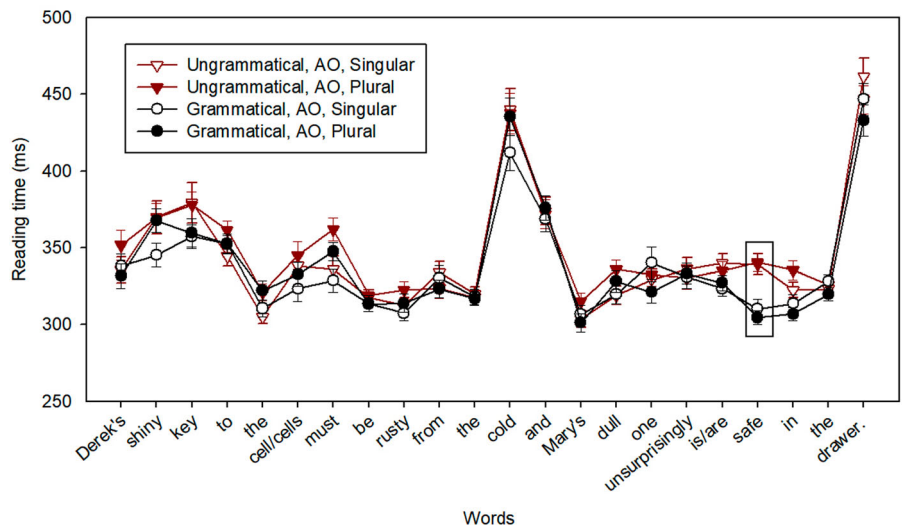
Figure 11. Region-by-region reading time means from Experiment 3b for Anaphoric *one* conditions. Error bars indicate the standard error. The regions of interest are is/are (verb), safe (spillover 1), and in (spillover 2).

At the critical verb region, a main effect of grammaticality was observed such that ungrammatical items were read slower than grammatical ones. A main effect of Anaphoric *one* was also observed such that items containing Anaphoric *one* were read slower than the baseline conditions. No main effects of local noun were observed, nor were any interactions between any factors.

At the verb spillover region 1, a main effect of grammaticality was observed such that ungrammatical items were read slower than grammatical ones. A main effect of Anaphoric *one* was again observed such that items containing Anaphoric *one* were read slower than the baseline conditions. Further subset analysis showed that this was driven by a main effect of Anaphoric *one* ($\beta = 18.54$, $SE = 6.10$, $t = 3.04$, $p < 0.01$) in ungrammatical

conditions, with no significant effect in grammatical conditions.

To contrast Anaphoric *one* with the baseline conditions, further subset analyses were performed. These revealed a main effect of local noun ($\beta = -11.72$, $SE = 5.57$, $t = -2.11$, $p < 0.05$) and grammaticality ($\beta = 17.34$, $SE = 5.76$, $t = 3.01$, $p < 0.01$) in the baseline conditions, and only a main effect of grammaticality ($\beta = 29.73$, $SE = 6.88$, $t = 4.32$, $p < 0.001$) in Anaphoric *one* conditions. We further conducted between-subject analysis where NPE (Experiment 1) and anaphoric *one* (Experiment 3) were directly compared.[7] At the spill-over region 1, there was a significant three-way interaction between Construction Type (NPE/Anaphoric *one*), Grammaticality and a Local Noun ($\beta = -34.38$, $SE = 15.92$, $t = -2.16$, $p <$
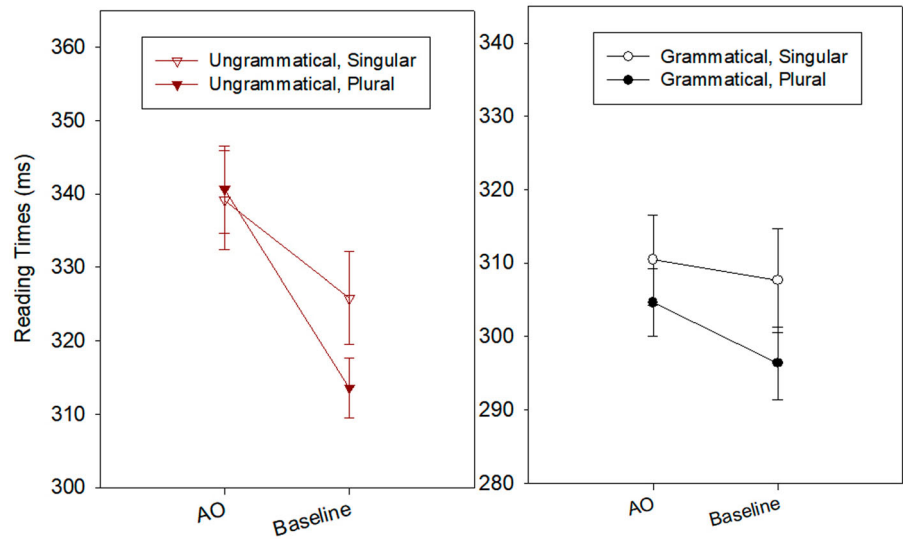

Figure 12. Reading times at the spillover region (*safe*) for all conditions in Experiment 3b. Error bars indicate the standard error.

Table 12. Summary of results of linear mixed effects models by region in Experiment 3b.

	Estimate	SE	t	p
Verb Region (*is/are*)				
by-subject random intercepts and slopes for Anaphoric *one*, and by-item random intercepts				
(Intercept)	322.92	7.46	43.29	
Local Noun	−2.87	3.55	−0.81	0.42
Grammaticality	8.79	3.55	2.47	0.05*
One	11.99	4.42	2.71	<0.001**
Local Noun × Grammaticality	−5.22	7.11	−0.73	0.46
Grammaticality × *One*	2.68	7.11	0.38	0.71
Local Noun × *One*	2.69	7.11	0.38	0.71
Local Noun × Grammaticality × *One*	−2.84	14.21	−0.20	0.84
Verb Region Spill-Over Region (*safe*)				
by-subject random intercepts and slopes for Local Noun, Grammaticality, Anaphoric *one*, Local Noun × Grammaticality, and Local Noun × Anaphoric *one*, by-item random intercepts and slopes for Local Noun, Grammaticality, and Anaphoric *one*.				
(Intercept)	316.13	8.44	37.47	
Local Noun	−6.86	4.57	−1.50	0.14
Grammaticality	23.58	5.07	4.65	<0.001***
One	12.28	4.35	2.82	<0.01**
Local Noun × Grammaticality	3.95	9.19	0.43	0.67
Grammaticality × *One*	12.64	6.80	1.86	0.06
Local Noun × *One*	9.31	7.63	1.22	0.22
Local Noun × Grammaticality × *One*	7.54	13.61	0.55	0.58
Verb Spill-over Region 2 (*in*)				
by-subject random slopes, and by-subject intercepts for Local Noun, Grammaticality, Anaphoric *one*, Grammaticality × Anaphoric *one*, and Local Noun × Anaphoric *one* and by-item random slopes, and by-item intercepts for Local Noun, Grammaticality, and Anaphoric *one*				
(Intercept)	317.17	7.14	44.40	
Local Noun	1.28	3.98	0.32	0.75
Grammaticality	13.51	4.23	3.19	<0.01**
One	2.71	4.13	0.66	0.51
Local Noun × Grammaticality	8.69	6.22	1.40	0.16
Grammaticality × *One*	6.36	7.57	0.84	0.40
Local Noun × *One*	1.30	6.65	0.20	0.84
Local Noun × Grammaticality × *One*	16.70	12.44	1.34	0.18

*** $p < .001$. ** $p < .01$. * $p < .05$.

0.05). There was also a main effect of Grammaticality ($\beta = 18.05$, $SE = 4.72$, $t = 3.83$, $p < 0.001$) and an interaction between Grammaticality and Construction Type ($\beta = -28.97$, $SE = 8.51$, $t = -3.40$, $p < 0.001$). This provides further support that NPE and anaphoric *one* behave differently.

At the verb spillover region 2, a main effect of grammaticality was observed such that ungrammatical sentences were read slower than grammatical sentences. Neither the main analysis nor any subset analysis revealed any main effects of local noun or Anaphoric *one* in either grammatical or ungrammatical conditions.

5.2.5. Discussion (Experiment 3a & 3b)

Experiment 3a and 3b aimed to further investigate how the parsers" sensitivity to grammatical distinction impacts processing of anaphoric elements, replacing the NPE in Experiment 1a and 1b with anaphoric *one* to test whether the retrieval of NPE involves accessing an antecedent without making a distinction between the head and the modifier. Similar to NPE, anaphoric

one should trigger the search for an antecedent, where the parser distinguishes the head noun and modifier. In contrast to NPE, anaphoric *one* relies heavily on a morphological cue to readily refer to its antecedent in memory. Thus, when the parser finds an antecedent that mismatches the number feature of the verb, it may filter out the local noun as a candidate. This would lead to the lack of agreement attraction for anaphoric *one*.

Note, further, although anaphoric *one* needs to access and reactivate the antecedent, given its nature as a pronominal (deep anaphora), it does not require the linguistic antecedent (Hankamer & Sag, 1976). Thus, it is possible that the parser does not build the structure of the antecedent when anaphoric *one* is encountered but rather finds its semantic or referential antecedent in the discourse representation.

Results of an offline acceptability judgment task (Experiment 3a) showed an overall interaction between local noun number and verb grammaticality. However, pairwise comparisons revealed that this difference was driven by the baseline condition only, with no attraction effects to items containing anaphoric *one*. Results of an online processing (Experiment 3b) were similar, also revealing no attraction effects in ungrammatical conditions containing anaphoric *one*.

This pattern suggests that cues like agreement features are potentially retrieved at the initial stage of parsing. Since there are multiple aspects of the head that match the retrieval cues associated with *one* (e.g. singular NP and noun category), the parser may select the head as a plausible subject, obviating an additional memory retrieval to access another element in the antecedent (e.g. local noun). This means that when processing anaphoric *one*, the parser puts the priority on the head noun over the local noun.

When the parser accesses a verb that matches the head noun, agreement is successful at first pass. However, in the number mismatching case, the parser only accesses the head noun and disregards the local noun. Accordingly, the parser does not need to undergo reanalysis because the head matches the morphological content of the retrieval cue, allowing it to more reliably access the head of the antecedent. The lack of agreement attraction in the Anaphoric *one* condition can therefore be attributed to the fact that the subject head noun matches multiple cues of the retrieval cue, making it unsusceptible to further interference effects. Furthermore, Anaphoric *one* may find its referential antecedent in the discourse representation as a deep anaphora (Hankamer & Sag, 1976). The lack of agreement attraction in Anaphoric *one* suggests that NPE and anaphoric *one* access antecedents differently, and that in

contrast to anaphoric *one,* NPE involves the recovery of the antecedent within the NPE-site.

Note that the lack of the agreement attraction effect could also be due to the intrinsic property of Anaphoric *one* itself. For the sake of exposition, let us walk through a speculative time course of the processing of Anaphoric *one.* First, when Anaphoric *one* is recognised, the parser can access and reactivate the antecedent of Anaphoric *one.*

(8)   Derek's key to the cells must be rusty from the cold and Mary's dull **one** ...

When the antecedent is reactivated, the number agreement of Anaphoric *one* and the antecedent should be inspected. Since Anaphoric *one* carries explicit morphological marking (e.g. *one* vs. *ones*), when the number marking of the Anaphoric *one* and the antecedent mismatch, the prediction is that items in which the antecedent and anaphor mismatch in number will elicit a reading time slowdown at the Anaphoric *one* site; in the present experiments, the head noun of the antecedent NP and the anaphor were both always singular, leaving this as an open question for future work.[8] Next, when the verb is encountered, agreement of the verb and Anaphoric *one* should be inspected because an overt noun with an explicit number marking is found in the subject position, much like in the non-ellipsis baseline conditions. Here, if the number marking of the verb and Anaphoric *one* mismatch, then such mismatch should give rise to slower reading of the verb.

(9) ... and Mary's dull **one** unsurprisingly <u>are</u> ...

Once the agreement mismatch is recognised, and if the antecedent of Anaphoric *one* is reactivated, then it is possible that the parser finds the plural local noun in the antecedent NP, and the verb can be erroneously licensed by the plural local noun in the antecedent.

(10) Derek's key to the <u>cells</u> ... and Mary's dull **one** unsurprisingly <u>are</u>

However, because Anaphoric *one* is explicitly number marked, and because the number agreement is inspected between Anaphoric *one* itself and the verb, the effect of erroneous licensing can be masked and not detectable.

Note further that, although there was no overall interaction between the local noun and grammaticality, the baseline conditions patterned broadly like previous experiments. The lack of agreement attraction in Experiment 3b might reflect an experimental artefact, namely that many trials contain an anaphoric *one* and a singular head. Given that the parser disregards the local noun if the head noun and the verb does not

match in the anaphoric *one* context, future work might investigate how the parser behaves if the head noun of the antecedent and the verb are both plural, serving to change the context of the ungrammaticality and possibly eliciting novel patterns of agreement attraction.

6. General discussion

This series of studies aimed to reveal whether grammatical information elided by NPE constrains the retrieval of the antecedent. We sought to investigate what kind of information is retrieved in NPE and other types of nominal anaphora constructions, testing structure retrieval in varying conditions with offline and online methodologies. In all six experiments, we took advantage of agreement attraction, the finding that the processing cost of ungrammatical verbs is attenuated by the presence of a feature matching intervenor.

In Experiment 1a and 1b, we examined acceptability judgments and processing of sentences containing NPE contrasted with sentences containing overt NPs (the baseline), with the aim to understand whether grammatical information is retrieved at the NPE site. The results showed that verb-matching local NPs provide an illusion of grammaticality and this illusion occurs in the NPE context as it does in the baseline conditions. Attraction was not observed in grammatical conditions in either NPE or the baseline, which constitutes further evidence for an asymmetry in agreement attraction (Lago et al., 2015; Tanner et al., 2014; Wagers et al., 2009). In terms of the retrieval mechanism, the implication is that when the features of the verb mismatch what the parser predicts, cue-based retrieval is recruited to fix the detected number disagreement (Lago et al., 2015; Parker & Phillips, 2017; Wagers et al., 2009). This is why agreement attraction was observed in ungrammatical but not grammatical conditions; in grammatical conditions, the calculation of agreement is successful on the first pass, thus the parser does not need to fix the number violation.

A plausible alternative account of the results from Experiments 1a and 1b is that the conjoined phrases serve to cue a parallel structure, which would require reactivating the elements in the first conjunct (Arregui et al., 2006; Callahan et al., 2010; Dickey & Bunger, 2011; Frazier et al., 2000; Frazier & Clifton, 2001; Kehler, 2000; Poirier et al., 2010; Shapiro et al., 2003; Sturt et al., 2010; Tanenhaus & Carlson, 1990). Thus, the presence of a conjoined phrase could trigger the parser to retrieve elements of the first conjunct without necessarily retrieving the antecedent itself. Experiment 2a and 2b were designed to test this alternative hypothesis by adding a No Anaphora condition that replaced the

noun in the first conjunct with a new noun, meaning that there was no anaphoric element for the parser to access and retrieve in the first conjunct. If coordination itself triggers the retrieval of the elements within the first conjunct, then we would expect agreement attraction even when the second conjunct contains an entirely new noun. The results of Experiment 2a and 2b do not support this alternative hypothesis; agreement attraction was not observed in reading times or acceptability judgements when NPE was replaced with a No Anaphora condition. This further suggests that the parser is not merely assessing the information in the first conjunct to search for the matching plural noun in the left-context.

Another alternative hypothesis is that the parser reactivates some information about the antecedent without distinguishing between the head and the modifier. To address this possibility, Experiments 3a and 3b examined constructions involving anaphoric *one,* which, like NPE, needs to access and reactivate the antecedent.

We predicted that if NPE can refer to an antecedent without the sensitivity to the grammatical properties, then NPE and anaphoric *one* should elicit similar agreement attraction effects. However, our results stood against this, showing no agreement attraction effects for anaphoric *one*. As both NPE and anaphoric constructions are similar in that they both need to access and reactivate the antecedent, this difference suggests that the way the antecedent is accessed in anaphoric *one* must be different from NPE. We suggest that processing NPE requires retrieval of grammatical information at the NPE-site, unlike anaphoric *one*. Anaphoric *one* is a pronominal, anaphoric, element. Thus, its interpretation is dependent on its antecedent. However, Anaphoric *one*, as a deep anaphora, does not require a linguistic antecedent (Hankamer & Sag, 1976). Therefore, how the antecedent is represented for anaphoric *one* can be different from NPE, namely that anaphoric *one* requires semantics and discourse related information of the antecedent, rather than grammatical information of the antecedent.

As such, our observed data have several implications for the structure and processing of NPE. The interpretation of the NPE-site is dependent on the antecedent NP ([$_{DP}$ *Derek's* [$_{NP}$ *key to the boxes*]]). Thus when the parser recovers the content of the NPE-site, the parser needs to access the information of this antecedent NP. As outlined in the introduction, the parser could use a variety of cues to do so, using case, category, animacy, number, and so forth to recover the content of the ellipsis site. It is plausible that the parser might only retrieve information of the head noun (e.g. *key*) because it is clear that the head noun is missing in the NPE site (see Dillon et al., 2013 for related discussion). The head noun shares

several features that match the element that is missing, namely the category *noun*, and meaning, *key*. The head noun is also the locus of the main meaning of the whole NP, making it the most prominent element within the NP. However, retrieving only the head noun of the antecedent NP would elicit no agreement attraction, as there is no local noun to attract the verb. Our data rule out this account, as we observed robust agreement attraction in NPE contexts.

It is also plausible that features associated with the head noun and the modifier would be accessed and retrieved simultaneously. The syntactic and morphological features borne out by both nouns would be at play and the parser would not necessarily privilege the head over the local noun, as the features of the local noun are equally accessible and similar to the features of the head noun. Therefore, we would expect an agreement attraction effect across NPE conditions (with grammatical and ungrammatical verbs) and in baseline contexts.

As a whole, our results support the idea that when processing the ellipsis site, the parser uses grammatical information. In other words, antecedent retrieval process involves recovering grammatical information at the initial stage of processing. When the parser encounters the genitive NP (*Mary's*) located at the beginning of the clause as well as an adverb (*unsurprisingly*), it is able to recognise the presence of the ellipsis site. When the NPE-site is processed, the parser is then able to access and retrieve the antecedent. The verb's agreement morphology can be predicted if the parser retrieves the number feature of the head noun of the antecedent. The head noun and the entire antecedent predict an upcoming singular verb; when this is violated by an ungrammatical plural verb, the modifier can trigger attraction in NPE.

Our data suggest that when the NPE-site is recognised, the parser carries out the following processes: (i) the parser retrieves the information associated with the head of the antecedent NP, ([[$_{head-N}$ *key*]], (ii) calculates the agreement between the head and the verb, and (iii) when the verb and the head noun do not have number agreement, the parser appeals to content-addressable memory and starts looking for another noun that could agree with the verb.

Retrieval of the head and modifiers results in agreement attraction in ungrammatical verbs following NPE. That is, if retrieval is triggered upon recognising the plural noun paired with an ungrammatical verb, then whatever plural noun in the left context should be accessed only if the head noun is recovered into the ellipsis site. Thus, our results show that the parser retrieves grammatical properties associated with the

NP from memory which is then used to construct the elided NP at the NPE-site.

The comparison of online and offline results across a variety of ellipsis configurations allows this study to provide unique insight into the timing of number mismatch detection, grammaticality effects and agreement attraction which our data suggest differ between ellipsis, overt NPs and other nominal anaphora constructions. In the NPE experiments (Experiments 1a and 1b), unlike previous research (Lago et al., 2015; Parker & Phillips, 2017; Tanner et al., 2014) and the No Anaphora and Anaphoric *one* experiments (Experiments 2a–3b), we observed no grammaticality effect prior to the agreement attraction. Instead, the grammaticality effect appeared simultaneously with agreement attraction, suggesting that the effect of the verb was observed after the retrieval of the elided element.

We suggest this difference in time profiles might be attributed to the availability of morphological cues. For NPE, the parser can recognise the ellipsis site when the parser encounters the possessive marked noun and an adverb. Spelling out a possible time course of the recognition of the NPE-site, it should be like the following. When the parser encounters the possessive noun, e.g. "John's", the parser anticipates a noun head. Immediately after the possessor marked noun, there is an adverb. An adverb is grammatically not compatible with a NP (e.g. *John's terribly destruction of the table*), and thus upon encountering an adverb the parser recognises that the anticipation is failed, and also recognise the grammatical incompatibility between the NP and an adverb. This recognition of the grammatical incompatibility between the NP and an adverb leads to an reanalysis of the structure from the anticipated NP structure ([John's [NP]]) to the structure of sentence which involves NP and VP ([S [NP John's [NP]] [VP [Adv unsurprisingly] [VP …]]]). As a result of this reanalysis process, the parser recognises the missing NP, the NPE site. The recognition of the NPE-site, triggered by reanalysis, should thus engender substantial processing complexity, potentially masking the grammaticality effect at the verb region. NPE and other nominal anaphoric constructions require accessing the antecedent and recovering information from memory.

In case of other nominal anaphoric constructions, there are always overt nouns. The presence of overt nouns does not lead to the reanalysis and also they provide clear morphological cues which indicate the specific type of nouns in the antecedent. In addition, the absence of agreement attraction in anaphoric *one* can also be accounted for in terms of its superior cue reliability. Cue reliability in morphonological information has been proved to be a strong factor in that reliable marking blocks agreement attraction (Franck et al.,

2008; Hartsuiker et al., 2003; Vigliocco et al., 1995; Vigliocco & Zilli, 1999, among others). In case of NPE, because the NP is missing, there is no reliable morphological cue. In the case of anaphoric *one*, an overt pronominal *one* provides reliable marking for a singular noun.

However, the ellipsis site in the NPE context does not have morphological cues, as it is silent. The lack of morphological cues may make the recovery of the antecedent difficult in the processing of the NPE-site compared to other cases of nominal anaphora. Therefore, the implication is that the relatively late grammaticality effect on NPE compared to other nominal constructions arises because antecedent retrieval in this construction is not guided by morphology, making it harder for the parser to find an antecedent.

7. Conclusion

The current studies investigated the processing of NPE by contrasting the elicitation of agreement attraction in NPE, conjoined but non-anaphoric elements and anaphoric *one*. The results showed equivalent patterns of agreement attraction in ungrammatical NPE items and overt sentences, comparable to what has been observed previous work. These results suggest the parser prioritises and retrieves the head at the initial stage of processing and retrieves the local noun only when it is necessary in parsing NPE. The results of the time-course profiles from these constructions further suggest that the parser is sensitive to grammatical distinctions at the ellipsis site.

Notes

1. In our study, the condition is called "ungrammatical", but we do not mean that the mismatch between the antecedent site and the ellipsis site in terms of syntactic structure is ungrammatical. In the literature, it has been observed in many places that such mismatch is possible (Arregui et al., 2006; Frazier, 2008; Kim, Kobele, Runner, & Hale, 2011). Rather, by "ungrammatical", we intend that the number mismatch between the "retrieved" antecedent and the verb is ungrammatical. For example, when the antecedent which has the singular noun does not match in number with the subsequent verb, this situation is very similar to the ungrammatical conditions in non-ellipsis baseline conditions, where the head noun does not match in number with the subsequent verb. Because we are calling such conditions in the non-ellipsis baseline conditions, ungrammatical conditions, we are calling the comparable conditions in the ellipsis conditions, "ungrammatical" conditions.
2. As an anonymous reviewer pointed out, an assumption of linear mixed-effects model is that the residuals should be normally distributed. Residuals were

distributed symmetrically around zero, suggesting normality (Min = −3.14; Median = 0.02; Max = 3.02). Following a reviewer's suggestion, we also carried out a cumulative logit model (also known as proportional odds model) for Experiment 1a. This revealed similar results to the linear models reported below, with a significant main effect of NPE ($\beta = 0.60$, $SE = 0.09$, $z = 6.34$, $p < 0.001$), Grammaticality ($\beta = -1.00 = 0.10$, $t = -10.51$, $p < 0.001$), a significant interaction between Local noun number and Grammaticality ($\beta = 0.60$, $SE = 0.19$, $z = 3.23$, $p < 0.01$) and a significant interaction between NPE and Grammaticality ($\beta = 0.65$, $SE = 0.19$, $z = 3.49$, $p < 0.001$).

3. There is an alternative account with regards to whether the whole structure is retrieved at the NPE site. In cases where the head is initially retrieved, it is possible that the parser accesses the head and calculates agreement at the verb. If the number mismatch between the head and the verb arises, the cue-based retrieval mechanism is employed. Even in this scenario, the parser is sensitive to the structural information such as the head and the modifier. Thus, the parser privileges the head noun over the local noun in the modifier, using structural information. In other words, the parser distinguishes the head and the modifier when it accesses an antecedent. At this point, it is hard to tease apart whether the whole structure or the head noun is retrieved at the initial stage of the retrieval processes.

4. Given that coordinate structures were used in these experiments, an expectation of parallelism could have led to easier retrieval of the head, weakening the potential interference from the modifier.

5. Again, residuals followed a symmetrical distribution around zero, suggesting normality (Min = −3.15; Median = −0.02; Max = 3.99). As in Experiment 1a, following a reviewer's suggestion, we also carried out a cumulative logit model (also known as proportional odds model) of Experiment 1b. In this analysis, we found significant main effects of Local noun number ($\beta = 0.24$, $SE = 0.08$, $z = 2.87$, $p < 0.01$), NPE ($\beta = -0.25$, $SE = 0.08$, $t = -3.04$, $p < 0.01$) and Grammaticality ($\beta = -2.53$, $SE = 0.10$, $z = -25.80$, $p < 0.001$), a significant interaction between Local noun number and Grammaticality ($\beta = 0.52$, $SE = 0.17$, $z = 3.12$, $p < 0.01$) and no other interactions.

6. Residuals followed a symmetrical distribution around zero, suggesting normality (Min = −4.53; Median = 0.06; Max = 3.12). As in Experiments 1a and 2a, we also carried out a cumulative logit model (also known as proportional odds model) of Experiment 3a. This disclosed main effects of Anaphoric *one* ($\beta = -0.26$, $SE = 0.08$, $t = -3.16$, $p < 0.01$) and Grammaticality ($\beta = -1.99$, $SE = 0.09$, $z = -21.36$, $p < 0.001$) and an interaction between Local noun number and Grammaticality ($\beta = 0.59$, $SE = 0.17$, $z = 3.48$, $p < 0.001$). There was also a marginal interaction between Local noun number, Grammaticality and Anaphoric *one* ($\beta = -0.65$, $SE = 0.34$, $z = -1.92$, $p = 0.06$).

7. This was suggested by an anonymous reviewer. We are grateful to the reviewer for pointing out this possibility.

8. Note that, "ungrammatical" conditions are called "ungrammatical" conditions because the number marking of Anaphoric *one* and the verb are not matched. They are ungrammatical not because the number marking of *one* and the antecedent mismatches, but because number marking of *one* and the verb mismatches.

References

Acuña-Fariña, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua. International Review of General Linguistics. Revue internationale De Linguistique Generale*, *143*, 108–128.

Arregui, A., Clifton, C., Frazier, L., & Moulton, K. (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language*, *55*(2), 232–246. doi:10.1016/j.jml.2006.02.005

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-Effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep It maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4, https://cran.r-project.org/web/packages/lme4/index.html. *R package version*

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

Clifton, C. J., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Italian Journal of Linguistics*, *11*(1), 11–39.

Crain, S. (1994). *Language acquisition in the absence of experience*. Cambridge, MA: MIT Press.

Dickey, M. W., & Bunger, A. C. (2011). Comprehension of elided structure: Evidence from sluicing. *Language and Cognitive Processes*, *26*(1), 63–78. doi:10.1080/01690961003691074

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*(2), 85–103. doi:10.1016/j.jml.2013.04.003

Drummond, A. (2011). Ibex Farm, Online server: spellout.net/ibexfarm

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, *56*(3), 357–383. doi:10.1016/j.jml.2006.07.004

Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, *23*(3), 329–374.

Frazier, L. (2008). Processing ellipsis: A processing solution to the undergeneration problem? In C. B. Chang, & H. J. Haynie (Eds.), *Proceedings of the 26th west coast conference on formal linguistics* (pp. 21–32). Somerville, MA: Cascadilla Proceedings Project.

Frazier, L., & Clifton, C. J. (2001). Parsing coordinates and ellipsis: Copy α. *Syntax (Oxford, England)*, *4*(1), 1–22. doi:10.1111/1467-9612.00034

Frazier, L., Munn, A., & Clifton, C. (2000). Processing coordinate structures. *Journal of Psycholinguistic Research*, *29*(4), 343–370. doi:10.1023/a:1005156427600

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*(6), 1411–1423. doi:10.1037//0278-7393.27.6.1411

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, *51*(1), 97–114. doi:10.1016/j.jml.2004.02.003

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(6), 1304–1321. doi:10.1037/0278-7393.32.6.1304

Hankamer, J., & Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry*, *7*, 391–426.

Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & Cognition*, *31*(8), 1316–1326.

Hornstein, N., & Lightfoot, D. (1981). *Explanation in linguistics: The logical problem of language acquisition*. London; New York: Longman.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or Not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. doi:10.1016/j.jml.2007.11.007

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238. doi:10.1037/0096-3445.111.2.228

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, *56*(3), 384–409. doi:10.1016/j.jml.2006.09.003

Kehler, A. (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy*, *23*(6), 533–575. doi:10.1023/a:1005677819813

Kim, C. S., Kobele, G. M., Runner, J. T., & Hale, J. T. (2011). The acceptability cline in VP-ellipsis. *Syntax (oxford, England)*, *14*, 318–354.

Kush, D. (2013). Respecting relations: Memory access and antecedent retrieval in incremental sentence processing. In C. Phillips, N. Hornstein, M. Israel, E. Lau, & J. Lidz (Eds.): ProQuest Dissertations Publishing.

Kush, D., & Phillips, C. (2014). Local anaphor licensing in an SOV language: Implications for retrieval strategies. *Frontiers in Psychology*, *5*, 1252. doi:10.3389/fpsyg.2014.01252

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in spanish comprehension. *Journal of Memory and Language*, *82*, 133–149. doi:10.1016/j.jml.2015.02.002

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, *25*(1), 93–115. doi:10.1007/bf01708421

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

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. doi:10.1016/j.tics.2006.08.007

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at Age 18 months. *Cognition: International Journal of Cognitive Science*, *89*(3), 295–303. doi:10.1016/S0010-0277(03)00116-1

Lightfoot, D. (1989). The child's trigger experience: Degree-0 learnability. *Behavioral and Brain Sciences*, *12*(2), 321–334.

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

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of experimental psychology: Learning. Memory and Cognition*, *35*(5), 1231–1239. doi:10.1037/a0016271

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: Evidence from sluicing. *Journal of Memory and Language*, *64*(4), 327–343. doi:10.1016/j.jml.2010.12.006

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, *29*(2), 111–123. doi:10.1023/A:1005184709695

McElree, B., & Dosher, B. A. (1989). Serial position and Set size in short-term-memory - the time course of recognition. *Journal of Experimental Psychology: General*, *118*(4), 346–373. doi:10.1037//0096-3445.118.4.346

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and*

Language, *48*(1), 67–91. Pii S0749-596x(02)00515-6. doi:10.1016/S0749-596x(02)00515-6

Nicol, J., Forster, K. I., & Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal of Memory and Language*, *36*(4), 569–587. doi:10.1006/jmla.1996.2497

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language*, *94*, 272–290. doi:10.1016/j.jml.2017.01.002

Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology*, *69*(5), 950–971. doi:10.1080/17470218.2014.992445

Payne, J., Pullum, G. K., Scholz, B. C., & Berlage, E. (2013). Anaphoric one and its implications. *Language: Journal of the Linguistic Society of America*, *89*(4), 794–829. doi:10.1353/lan.2013.0071

Pearl, L., & Lidz, J. (2013). Parameters in language acquisition. The Cambridge handbook of biolinguistics. In K. Grohmann, & C. Boeckx (Eds.), *The Cambridge handbook of biolinguistics*. Cambridge: Cambridge University Press.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*(3), 427–456. doi:10.1006/jmla.1999.2653

Phillips, C., Wagers, M., & Lau, E. F. (2011). Grammatical illusions adn selective fallibility in real-time language comprehension. *Syntax and Semantics 37: Experiments at the Interfaces*, *37* (2011), 53–186. doi:10.1108/S0092-4563(2011)0000037009

Poirier, J., Wolfinger, K., Spellman, L., & Shapiro, L. P. (2010). The real-time processing of sluiced sentences. *Journal of Psycholinguistic Research*, *39*(5), 411–427.

Rohde, D. (2003). Linger: a flexible platform for language processing experiments, version 2.94, Online: http://tedlab.mit.edu/~dr/Linger

Shapiro, L. P., Hestvik, A., Lesan, L., & Garcia, A. R. (2003). Charting the time-course of VP-ellipsis sentence comprehension: Evidence for an initial and independent structural analysis. *Journal of Memory and Language*, *49*(1), 1–19. doi:10.1016/S0749-596X(03)00026-3

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562. doi:10.1016/S0749-596X(02)00536-3

Sturt, P., Keller, F., & Dubey, A. (2010). Syntactic priming in comprehension: Parallelism effects with and without coordination. *Journal of Memory and Language*, *62*(4), 333–351. doi:10.1016/j.jml.2010.01.001

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

Tanner, D., Grey, S., & van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology*, *54*(2), 248–259. doi:10.1111/psyp.12788

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, *76*, 195–215. doi:10.1016/j.jml.2014.07.003

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, *48*(4), 740–759. doi:10.1016/S0749-596x(03)00003-2

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(2), 407–430. doi:10.1037/0278-7393.33.2.407

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316. doi:10.1016/S0749-596x(03)00081-0

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166. doi:10.1016/j.jml.2006.03.007

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263. doi:10.1016/j.jml.2011.05.002

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, *34*(2), 186–215.

Vigliocco, G., & Franck, J. (1999). When Sex and syntax Go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, *40*(4), 455–478.

Vigliocco, G., & Zilli, T. (1999). Syntactic accuracy in sentence production: Gender disagreement in Italian language impaired and unimpaired speakers. *Journal of Psycholinguistic Research*, *28*, 623.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237. doi:10.1016/j.jml.2009.04.002