NORTHWESTERN UNIVERSITY

Hold, Release, and Retrieve: The Study of Wh-Filler-Gap Dependencies and Ellipsis

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By

Nayoun Kim

EVANSTON, ILLINOIS

September 2019

Abstract

This dissertation is concerned with how components in memory structures and online structure building processes interact by investigating the online processing of Wh-Filler-Gap Dependencies (WhFGD) and ellipsis constructions. Resolving long-distance dependencies involves linking the dependent element to the controlling element. In the case of Wh-gap dependency formation, the wh-element is linked to the gap. In the case of ellipsis resolution, the ellipsis site is linked to the antecedent. In the processing of long-distance dependency resolution, I point out that two component processes are involved: the storage/maintenance component and the retrieval component. A series of studies on WhFGD formation reveal that the sentence processing mechanism involves the maintenance component on top of the retrieval component. Studies on ellipsis constructions further reveal that when the antecedent is retrieved, detailed grammatical structural information should be retrieved, thus grammatical and structural information must be encoded in memory. Based on the results of these studies, I specifically argue for the following points: (i) the filler is released from memory, depending on the grammatical requirement of the filler; (ii) given that information associated with the filler being retrieved reflects the extent to which the filler is maintained, the parser retrieves grammatical information associated with the wh-filler; and (iii) the parser is sensitive to grammatical distinctions at the ellipsis site in contrast to the processing of Anaphoric *one* and Pronoun *it*. These studies provide evidence that both the maintenance and retrieval process are heavily constrained by grammatical information associated with the elements that engage in dependency formation.

Acknowledgment

My deepest gratitude goes to my advisor, Masaya Yoshida, and members of my committee: Laurel Brehm, Matt Goldrick, Colin Phillips, and Alexis Wellwood.

First and foremost, I would like to thank my advisor, Masaya Yoshida. Working with him has truly been an amazing experience in my life. He helped spark my interest in carrying out research at the intersection of psycholinguistics and syntax. Honestly speaking, there were times that I could not understand some of his random comments immediately, which only later returned as an 'aha (This is exciting! Did Masaya already knew this long before, while I was still focusing on the trees?)-moment'. Looking back, I am deeply grateful that he provided me with opportunities to explore and discover every direction by myself (although he was guiding me at every stage unnoticed). I appreciate his generous insightful comments on my half-baked ideas and analyses, while also valuing his desire for clearer and more cogent argumentation. His demands for precision have helped me investigate core interesting problems in my research. Thank you for supporting me in pursuing my passion and constantly reminding me to never be afraid of being wrong. It has been an honor for me to have you as my advisor and mentor in my life.

I would like to thank Laurel Brehm, Matt Goldrick, Colin Phillips, and Alexis Wellwood for serving on my committee and providing unwavering support and guidance. I would like to thank Laurel Brehm for guiding me in uncovering the underlying mechanisms behind statistics and data analyses, and for being a model of balancing professionalism with optimism and enthusiasm. Working with her has allowed me to view data from different perspectives in psycholinguistics and to become a more thorough psycholinguist. I thank Matt Goldrick for his unwavering support and encouragement, and for being a constant source of inspiration on critical and clear thinking and comprehensive data analyses. I am also thankful for his role as a chair in the Linguistics department for making such a stimulating and ideal environment to work on psycholinguistics. I would like to thank Colin Phillips who has contributed significantly to my research from the beginning of my career. He has played an unquestionable role in shaping my ideas about psycholinguistics, and this dissertation is largely inspired by his research. I am grateful for his encouragement and support as well as for pushing me to articulate and pursue exciting ideas. Alexis Wellwood has constantly asked fundamental and deep questions about my research at every stage of my graduate career. I am grateful for her generosity in time and support during her time at Northwestern, and for her influence on the ways in which I approach problems in cognitive science.

This dissertation has greatly profited from the guidance of other linguists outside Northwestern University. I would like to particularly thank Patrick Sturt, for his thoughtful and illuminating comments. His work on the intersection between syntax and psycholinguistics has inspired me tremendously, instilling in me a profound respect for combining theoretical works with rigorous experimental works. I would like to also thank Matt Wagers for reading substantial chapters of my dissertation and for his insightful comments, and for allowing me to formulate many of my research ideas. I would also like to thank Julie Van Dyke, who provided invaluable comments and support on my works.

Over the course of the dissertation projects, parts of this dissertation were presented and discussed in various venues. I would like to thank audiences at Center for Research in Language, Haskins Laboratories, Illinois Language and Literacy Initiative group at the Beckman Institute, Syntax, Semantics, and Sentence Processing Lab and Aphasia and Neurolinguistics Lab at Northwestern for their lively discussions and helpful comments. I

am also grateful to Keir Moulton and Daphna Heller for their belief in my potential; I am very excited for the next stage of my research career working with you two.

I would like to thank my teachers at Northwestern Linguistics Department. In particular, I would like to thank Ann Bradlow, Brady Clark, Jennifer Cole, Erin Leddon, Janet Pierrehumbert, and Gregory Ward, for enriching my ideas in linguistics through stimulating conversations and discussions. Janet Pierrehumbert deserves my special thanks. She was generous with her time and insights where I was able to synthesize and refine my research interests and ideas in my early years at Northwestern. I was able to rely on her unwavering support even after she left to Oxford.

I am also grateful to Cindy Thompson, who provided me with invaluable opportunities to work on projects in neurolinguistics, using different neuroscientific methods to illuminate syntax in our brain. I would also like to thank Robin Nusslock, professor in Psychology, whose office was right across my office in Swift Hall. He provided constant academic advice and encouragement in my graduate career, instilling me with determinism and enthusiasm.

I am thankful for my wonderful colleagues in Syntax, Semantics, and Sentence Processing Lab. I particularly thank Kathleen Hall, Shayne Sloggett, Ethan Myers, Wes Orth, Peter Baumann, and Mike Frazier for their stimulating discussions and comments on my works. I would like to also thank Yunyan Duan for late night lively discussions in our office, and her suggestions on data analyses. Thanks also goes to undergraduate research assistants: Alex Krauska, Jiayi Lu, Cecily Mejia, Anelia Kudin, Conor Jones, and Keeley Logsdon for their help in generating stimuli and help with running some of the experiments reported in the dissertation. Working with you all was a great joy and happiness to me.

My life at Northwestern would not have been as it is without my lovely cohorts. Many discussions regarding work and life in general with Kat Hall have been an indispensable source of memory during the times at Northwestern. I will always remember your family (Kat, Sean, Miles, and Mabel) when I think of PhD life. Tommy Denby has been a reliable and impeccable cohort, officemate, and a friend. I value his emotional and intellectual companionship, especially during our dissertation stages. Thank you for being such a genuine human being.

I would also like to thank our technician, Chun Liang Chan for his help in technical issues.

My deep gratitude goes to other linguists in and outside Northwestern: Brian Dillon, Luca Campanelli, Katy Carlson, Wing wee Chow, Kiel Christianson, Ian Cunnings, Mike Dickey, Cynthia Fisher, Kara Federmeier, Lyn Frazier, Grant Goodall, Jesse Harris, Matt Husband, Oriana Kilbourn-Ceron, Elsi Kaiser, Tyler Knowlton, Maryellen MacDonald, Maayan Keshev, Dave Kush, Sol Lago, Maryellen MacDonald, Shota Momma, Conny Myers, Jason Merchant, Dongwoo Park, Dan Parker, Pavel Logačev, Jeff Runner, John Trueswell, Matt Walenski, Eva Wittenberg, Ming Xiang, James Yoon, and Sion Yoon for their genuine support and many critical comments.

I thank my friends, Yoonji Kim, Midam Kim, Sujung Kim, Linyi Zhang, and Kyounghee Lee for their support and friendship.

I would also like to thank my teachers in Korea, to Hanjung Lee, Yongsoon Kang, Jihye Kim, and Nayoung Kwon for their warm encouragement and support over the past years. They triggered my interests in the investigations on linguistics as a window to better understand the human mind.

Parts of this dissertation have been graciously supported by NSF-DDRI grant. I thank NSF grant for their funding (NSF DDRI grant: BCS 1749580).

Table of Contents

List of Tables

List of Figures

1. General Introduction

Online sentence processing can be characterized as the dynamic interaction of incremental structure building and the various components of memory representations. This dissertation is concerned with how various components in memory structures and online structure building processes interact. To this end, this study investigates the syntax and the real-time processing of ellipsis constructions (versus other anaphoric elements), and Wh-Filler-Gap Dependencies (WhFGD), both of which crucially involve the process of structure building and memory retrieval mechanisms. This thesis also approaches the problem of active maintenance by investigating the interaction between grammatical properties of different wh-phrases and memory structures.

This thesis shows that studying ellipsis constructions and WhFGD constructions is particularly useful for understanding the mechanism working behind retrieval and maintenance. In ellipsis constructions, the interpretation of the ellipsis site is dependent on the antecedent. Thus, when the ellipsis site is processed, the parser needs to access the antecedent and retrieve the information associated with the antecedent. By investigating what information associated with the antecedent of ellipsis is retrieved or not when the ellipsis site is processed, we can observe what sort of information is accessible, as well as what sort of information is susceptible to memory decay. Similarly, investigating WhFGD with different types and qualities of fillers will inform us of how elements in a dependency relation are stored and maintained, what kind of information is maintained, and what information is released from memory and what has to be stored in memory again.

1.1. Introduction

This dissertation addresses the question of how our syntactic structures interact with various components of memory representations. Forming and resolving dependencies plays a crucial role in understanding human language. For instance, the grammatical status (the number feature) as well as the interpretation of the subject noun phrase (NP) is determined by the verb.

(1) a. The key to the cells is rusty from the cold.

b. * The key to the cells are rusty from the cold.

This demonstrates the property of the dependency formation: when the parser encounters the dependent element, the parser needs to search for the controlling element and link the controlling element to the dependent element: as holding the dependent element in memory is costly, the parser actively looks for the controlling element and tries to link the dependent element to the controlling element as soon as possible. In many constructions, the dependent element precedes the controlling element. In such cases, normally the dependent element can signal the presence of the dependency relation. Therefore, the parser can initiate a search for the controlling element upon encountering the dependent element. At the same time, it seems that the dependent element, which signals the presence of a dependency relation, is located after the controlling element; this suggests that in many cases, the parser can recognize the presence of a dependency relation and trigger the retrieval of a controlling element from memory only after the dependent element is encountered. Thus, depending on the positioning of the dependent element and the controlling element, there are two ways to form dependencies.

First, if the dependent element precedes the controlling element, then the dependent element will trigger the forward dependency formation. Given that the position of the controlling element can be farther along in the sentence and the exact position of the controlling element is often not signalled by the dependent element, the parser may not be able to resolve the dependency immediately. Because the dependent element cannot be interpreted and grammatically licensed before it is linked to the controlling element, the parser needs to hold the information of the dependent element in memory until it encounters the controlling element.

On the other hand, if the dependent element follows the controlling element, then the dependent element may trigger the backward dependency formation. In this process, the parser needs to recover the information of the already-processed element, i.e., the dependent element triggers the retrieval of the controlling element from memory. If the potential controlling element has already completed the dependency formation at the point when the dependent element is encountered, then the controlling element can be cached out from memory. If we assume that the already-processed materials are stored in the content-addressable memory store, then a parallel access retrieval process is triggered by the dependent element.

One of the aims of this study is to reveal the mechanism working behind the dependency formation process and how it interacts with memory representations by investigating the online processing of WhFGD formations (forward dependency formations) and ellipsis constructions (backward dependency formations)

The WhFGD formation process represents the proactive forward dependency formation process. The formation of WhFGD requires associating a wh-phrase with other elements that are grammatical licensors of a wh-phrase. An example of a WhFGD construction is illustrated in (2).

(2) Which mistake in the program will be harmful for everyone involved?

The grammatical status (the wh-phrase interpreted as the theme argument) and the meaning of the wh-phrase are dependent on other elements that are grammatical licensors of the wh-phrase (the controlling element). Although the wh-phrase signals that there is a dependency relation in which the gap should appear somewhere later in the sentence, the parser needs to hold the wh-phrase in memory until it is successfully linked to its licensor (the gap position).

In general, we can plausibly assume that this kind of proactive processing involves the following component processes: the parser recognizes the filler, stores the filler in memory, recognizes the gap, and forms the dependency. Once the dependent element is linked to the controlling element, the dependent element does not need to be stored in memory anymore, because the interpretation and the grammatical status of the dependent element can be determined upon forming the dependency. The parser can thus release or forget the filler once it forms the dependency. In other words, in addition to storing/maintaining the dependent element, there must be a process of releasing the dependent element from memory.

In contrast, let us examine one example of the backward dependency formation: the Noun Phrase Ellipsis (NPE) construction in which the meaning of the ellipsis site is dependent on the antecedent, but the nominal phrase in the second conjunct is missing.

(3) Derek's key to the cell must be safe in the drawer, and Mary's probably [$_{NP}$ ø] is rusty from the cold.

The missing part of the second conjunct in the DP (*Mary's*) should be dependent on the antecedent in the first conjunct (*key to the cell*) to fulfill the proper interpretation. This indicates that when the parser recognizes the NPE, the parser needs to access the antecedent and retrieve the associated information at the NPE-site.

This process represents a retroactive backward dependency formation process: the parser recognizes the ellipsis site and retrieves the antecedent. We show that the content of the retrieved element is different depending on what anaphoric element is processed.

Against this background, this dissertation aims to uncover how maintenance, release, and retrieval processes operate in ellipsis and WhFGD processing, by focusing on what information is retrieved at the ellipsis site and what information of the dependent element is maintained during the WhFGD formation.

In terms of maintenance, we contend that (i) the wh-filler should be maintained in memory and released once the wh-filler is linked to the controlling element; and (ii) if the wh-filler is released from maintenance and accessed or retrieved afterwards, it will be less accessible for the parser, and thus its retrieval will be less successful with the retrieval of coercer-grained information, compared to the cases where information is maintained over the course of the dependency formation. Specifically, we posit memory architecture where a working memory like special memory state is assumed in terms of the maintenance component. In terms of retrieval, we argue that different anaphoric elements give rise to different retrieval behaviours depending on whether the structural information associated with the antecedent is recovered or not. Specifically, we aim to show that the readers are able to retrieve fine-grained information as well as detailed syntactic structures associated with the antecedent, by utilizing grammatical/structural information as retrieval cues.

A series of studies on WhFGD formation reveal that the sentence processing mechanism involves a maintenance component on top of the retrieval component. Our studies show evidence that the maintenance component is working crucially in the forward dependency formation (Gibson, 1998). Sentence processing mechanism cannot be exclusively accounted for in terms of either the maintenance or retrieval, and theories on sentence processing should assume and incorporate such maintenance effects. How would this stack-like mechanism be combined to other memory dynamics (McElree, 2006)? The parser should work in both the storage (where the processed materials and retrieve-to-be materials are stored) as well as in the short-term storage like stack. Our data supports that maintenance is governed by grammatical information and that syntactic structures are accessed and utilized in working memory during online sentence processing.

In terms of the retrieval, previous studies have not been so clear whether grammatical/structural information can be utilized as retrieval cues (Kush, 2013; Parker, Shvartsman, & Van Dyke, 2017). However, our studies show that quite detailed syntactic structure is deployed as retrieval cues. Specifically, the results of our NP Ellipsis studies reveal that readers are able to retrieve fine-grained information as well as detailed syntactic structures at the verb position. This suggests that quite detailed syntactic structure is used as retrieval cues or otherwise it is hard to identify what the antecedent is, during the resolution of the antecedent retrieval at the NPE-site. Thus, these studies provide insights into what kinds of information could be considered as potential retrieval cues as well as what the theory of retrieval mechanisms should capture.

1.2. Outline of the Dissertation

1.2.1. Chapter 2: Maintenance and Retrieval

The goal of Chapter 2 is two-fold: (i) to explore differences in terms of the retrieved information at the gap location depending on the type and quality of the fillers, and (ii) to understand when different kinds of fillers are released from memory.

We show that the maintenance occurs in a proactive WhFGD formation. In the resolution of WhFGD, the parser actively searches for the gap so that the wh-filler can be released from memory and thus would no longer impact memory resources. We aim to uncover how maintenance and retrieval components in memory interact with the dependency formation. We present a series of experiments examining (i) how long the parser can hold the filler and (ii) which aspect of the fillers are maintained, and which are not maintained and need to be re-activated at a later stage.

First, we investigate how the maintenance component plays a role in the online WhFGD by examining the processing of different kinds of Wh-Phrases (WhPs: *who*, *how why)*. Syntactic studies on these WhPs have shown that *why* is linked to the TP (Ko, 2005; Yoshida, Nakao, & Ortega-Santos, 2015), *how* is linked to the VP, and *who* is linked to a verb (gap). This indicates that wh-phrases such as *who* and *how* need to be stored in memory until they are linked to their licensor, the verb, whereas *why* can be released from maintenance as soon as TP is recognized. Based on this, we argue that the differences in the storage cost effects depend on different types of wh-phrases, which result from the differences in the length of the WhFGD. These results provide strong evidence that the filler should be maintained in memory and released once the wh-filler is linked to the controlling element.

Based on our findings that maintenance[1] plays a role in the online dependency formation process, we ask the following question: how long can the parser hold the filler, and which aspects of the fillers are maintained relatively well? How do these characteristics influence the retrieval event? We use the agreement attraction phenomenon as a window into the information retrieval processes at the gap/verb position. If it is the case that the wh-filler needs to be maintained, then we expect to observe differences in terms of the retrieved information at the verb location depending on the type and quality of the fillers.

According to some theories (Nicenboim, Vasishth, Gattei, Sigman, & Kliegl, 2015), the retrieval mechanism should reactivate different fillers similarly; in other words, regardless of the type of the filler, it should be reactivated at the same point as any other. If different fillers are retrieved in a similar manner, we expect to observe no differences between the retrieved information for different types and qualities of fillers at the verb position. However, we observe instead that maintained elements and retrieved elements are processed differently, revealing different reactivation profiles. We show that depending on the filler types, the information that is retrieved at the point of the verb is different. More

---

[1] We use maintenance/storage in a sense that the element is stored in memory until the dependency is completed. Once the dependency is complete, information no longer exacts memory costs. When there is another dependency to be formed, the parser needs to reactivate again. This notion of maintenance is compatible with Gibson (1998)'s notion of maintenance/storage.

specifically, maintenance of information could lead to easier retrieval, allowing the parser to access richer and finer-grained information (category and grammatical information) but releasing the filler from memory only enables the parser to retrieve coarse-grained information of the filler. We hold that maintenance, release and retrieval should be considered during the formation and resolution of the WhFGD (Fiebach, Schlesewsky, & Friederici, 2002; Wagers & Phillips, 2014) as well as during parsing in general.

1.2.2. Chapter 3: Retrieval

The goal of Chapter 3 is to better understand the nature of the retrieval processes in ellipsis contexts. This chapter investigates further details about the retrieval mechanism. This chapter is concerned with how different kinds of anaphoric elements could result in different retrieval behaviors during real-time processing.

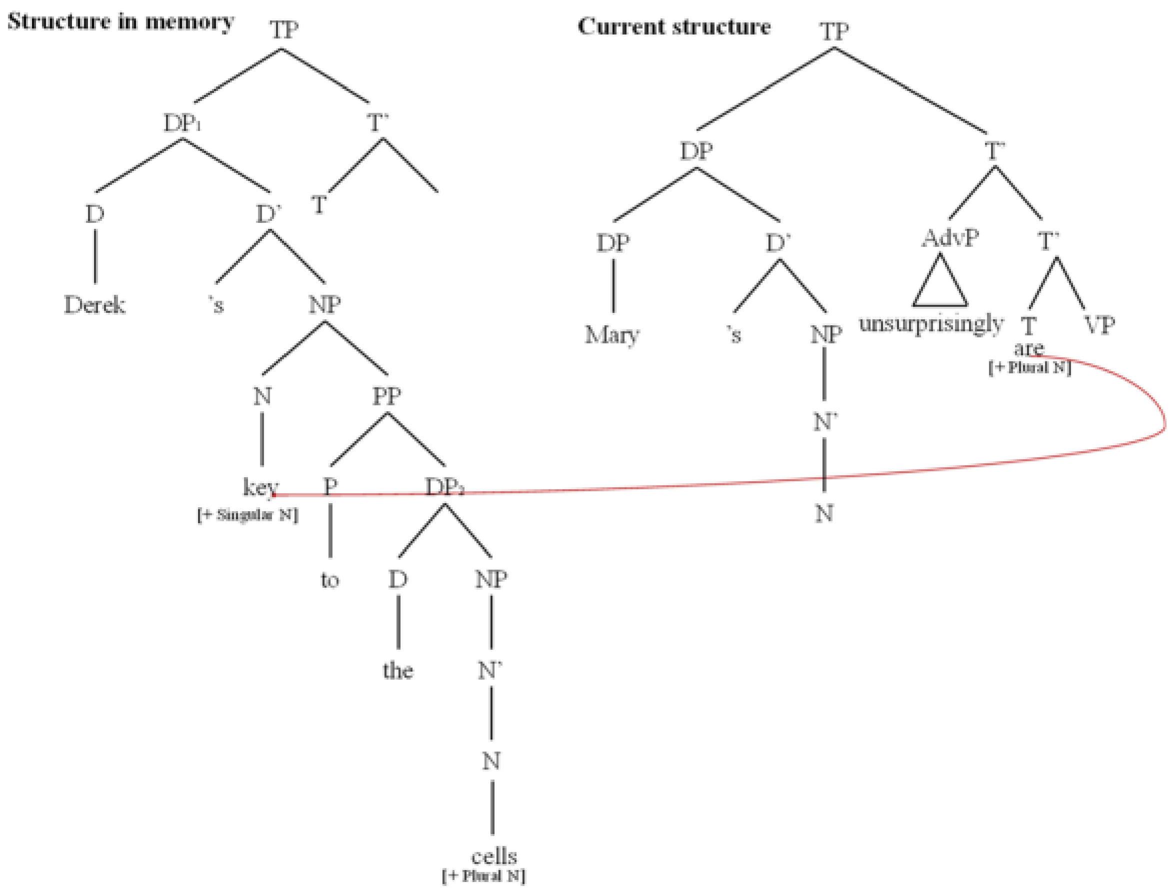
In order to examine the different retrieval behaviors of ellipsis and various anaphoric elements, we first examine the structure associated with the ellipsis site in the NPE construction. Thus, the first goal of chapter 3 is to demonstrate that NPE involves internal syntactic structure that stands in a certain parallel relation with the antecedent. This syntax component provides insights into the mechanisms underlying the antecedent retrieval process: the NPE-site is associated with a rich hierarchical structure, and therefore the recovery of the content of the ellipsis site should mean the recovery of the structure within the NPE-site. A detailed study of the syntax of NPE is needed in order to understand the processes of identifying and recovering the content of the NPE-site in regard to antecedent retrieval. Previous studies have suggested that the ellipsis-site of NPE (the NPE-site) is associated with certain pro-forms with no internal syntactic structures (Lobeck, 1995, 2007; Panagiotidis, 2003; Postal, 1969). In terms of NPE, there is apparent supporting evidence for a pro-form analysis, namely a ban on wh-extraction from the NPE-site. Contrary to this, we demonstrate that the NPE-site is associated with rich hierarchical structures by showing that NPE-sites exhibit properties characteristic of ellipsis constructions, such as vehicle change (Fiengo & May, 1994), inverse scope effects, and various connectivity effects.

Furthermore, we show that the restrictions on wh-extraction from the NPE-site follow from independently motivated requirements and conclude that NPE is surface anaphora. Specifically, we argue that wh-movement out of the NPE-site is blocked since the phase head (Chomsky, 2001) and the ellipsis-licensing head (Albrecht, 2010; Lobeck, 2007; Merchant, 2001, etc) are different. The NP that contains the launching site of wh-movement undergoes ellipsis before wh-movement within the DP takes place. That is, wh-movement out of the NPE-site is not possible because the NPE-site is licensed by a functional head generated lower than the phase head D, which attracts the overtly moving phrase. This predicts that in the context where D licenses the ellipsis of the nominal projection, wh-movement or A-bar extraction is possible.

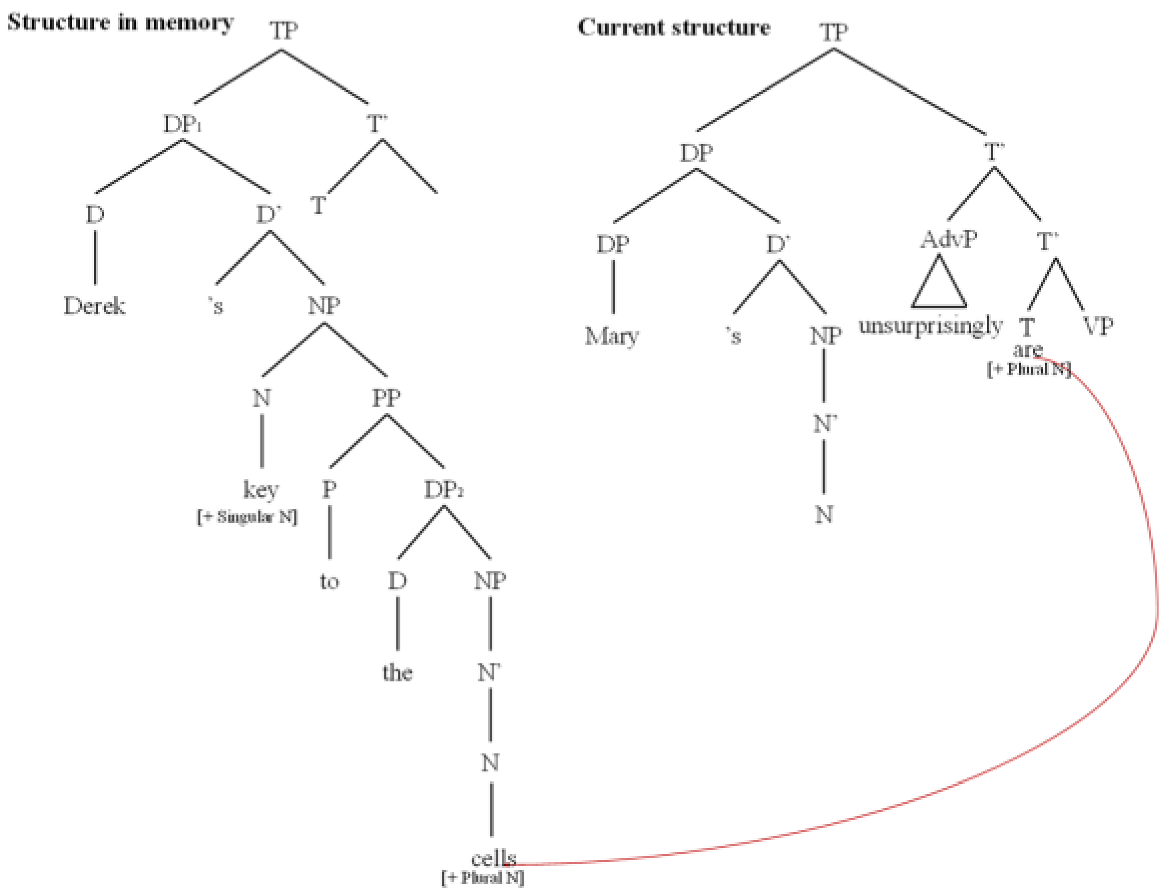
The detailed studies on the syntax of NPE help us further reveal the mechanism of antecedent retrieval working behind the processing of these different types of anaphoric constructions, and uncover the general mechanism of antecedent retrieval that forms the basis of the experimental psycholinguistic works addressed below. By using a grammaticality illusion paradigm observed in the typical agreement processing (Wagers, Lau, & Phillips, 2009), we show that information retrieved during ellipsis resolution differs from the information retrieved for non-elliptical nominal anaphora (e.g. Anaphoric *one* and pronominal Pronoun *it*).

The logic behind this paradigm is that if the grammatical information of the antecedent is retrieved, then the parser first calculates the agreement relation between the verb and the head noun of the retrieved antecedent, as shown in (4). When the agreement turns out to be ungrammatical (the head of the retrieved antecedent (e.g., *key*) does not match in number with the verb (i.e., *are*)), the parser retrieves the local noun to repair the agreement, which could lead to facilitation in the processing of the verb, as in (5).

(4)



(5)



The central findings of these studies are that when noun phrase ellipsis are processed, all the content – including grammatical information, such as the distinction between the head and the modifier within the antecedent NP – of the antecedent is retrieved and thus we observe an illusion of grammaticality. On the other hand, Anaphoric *one* and Pronoun *it,* which are pronominals, do not require a linguistic antecedent and thus the parser does not build a structure. This research raises questions about how different kinds of anaphoric elements give rise to different retrieval behaviors, which could in turn inform us about what linguistic information associated with the antecedent is accessed and retrieved when an anaphoric element is processed. By employing both offline and online experiments, our findings suggest that the grammatical differences of the anaphoric element (whether elliptical or not) control what element is retrieved.

We argue that both maintenance and retrieval components are crucial in accounting for the differences in different WhFGDs and ellipsis. We show that the maintenance device is crucial for storing structural information, as the dependent element needs to be usable for the dependency formation in some way or another. Further, readers are able to make use of fine-grained information such as detailed syntactic structure in the resolution of dependencies.

2. Maintenance and Retrieval[2]

2.1. Introduction

The study of WhFGD suggested that the mechanism working behind the online WhFGD formation may involve the following three component processes: storing the wh-filler in memory (Frazier, 1985; Gibson, 1998; Gibson & Thomas, 1999; Wagers & Phillips, 2009), searching for the gap or the licensor of the wh-filler (Aoshima, Phillips, & Weinberg, 2004; Crain & Fodor, 1985; Frazier & Flores D'Arcais, 1989; Lee, 2004; Omaki et al., 2015; Phillips, 2006; Pickering & Barry, 1991; Stowe, 1986; Traxler & Pickering, 1996), and linking the wh-filler to the licensor and reactivating the wh-filler in memory (Lewis &

[2] Portions of the Chapter 2 have been published in *Language, Cognition, and Neuroscience* (Kim, N., Brehm, L., Sturt, P., & Yoshida, M. (2019). How long can you hold the filler: maintenance and retrieval. *Language, Cognition and Neuroscience*, 1-26).

Vasishth, 2005; McElree, 2001, 2006; McElree & Dosher, 1989; Nicol, Fodor, & Swinney, 1994; Nicol & Swinney, 1989; Van Dyke, 2007; Van Dyke & McElree, 2006; Wagers & Phillips, 2014). However, recent studies on dependency processing have suggested that many of the effects associated with some of these component processes can be regarded as the effect of retrieval of the first element of the dependency when the second element of the dependency (e.g., the verb) is processed (Vasishth & Lewis, 2006). Against this background, this chapter investigates the maintenance/storage component of the online WhFGD formation processes. Through offline and online experiments, we attempt to show that the wh-filler is maintained/stored in memory during the processing of WhFGD sentences, and once the wh-filler is linked to the licensor, the wh-filler is released from memory, and thus forgotten.

Further, we attempt to reveal the mechanisms behind online WhFGD formation. Specifically, we aim to uncover how maintenance and retrieval operate in WhFGD processing, by paying special attention to what information is retrieved from the wh-filler when the gap is recognized. We contend that if the wh-filler is released from maintenance and retrieved at a later point, its activation in memory will be lower, and its retrieval will be less successful, relative to a situation where it is maintained.

2.1.1. Wh-Filler-Gap Dependency Formation

One of the prominent properties of human language is that filler-gap dependencies, exemplified by wh-interrogative constructions like (6), can potentially span across an unbounded number of words (unbounded dependency: Chomsky, 1977).

(6) What did John eat yesterday?

The WhFGD, in particular, crucially involve two elements, the wh-filler and, the gap, the verb or the preposition, where the interpretation and the grammatical status of the wh-filler is crucially dependent upon the gap. In other words, the grammatical status (e.g., whether the wh-element is the subject or the object, and whether the wh-element bear nominative case or accusative case) and the interpretation (e.g., whether the wh-element is interpreted as an agent or patient) of the wh-element are all determined in relation to the verb, the preposition, or the gap that the wh-element is linked to (Huddleston & Pullum, 2002: 1079, for clear illustrations of this type of dependency relations). For example, in (6), the wh-filler, *what*, is interpreted as the patient (something to be eaten), and its function is the direct object of the verb *eat*. Thus, the interpretation and grammatical status of *what* in (6) are crucially *dependent* on the verb *eat*. As such, if the object of *eat* is realized by another NP, e.g., *sushi* (**What did the student in the classroom eat sushi?*), the example is not acceptable because the wh-filler *what* cannot be interpreted. Thus, sometimes, the wh-filler is referred to as a *dependent element* and the verb, the preposition or the gap, which control the interpretation and the grammatical status of the wh-filler, is referred to as a *controlling element*.

This property of the WhFGD construction raises a question of how it can be processed and interpreted during incremental sentence processing. In WhFGD, the wh-element needs to be held in memory till it reaches the gap to achieve the correct interpretation. The resolution of WhFGD requires two processes: maintaining information until the gap, and retrieving information. However, in principle, the distance between the filler and the controlling element (e.g., the gap) can be potentially long. This motivates the parser to resolve the dependency as soon as possible (link to the controlling element) so that the memory costs are no longer taxed[3]. As Gibson (1998) argues, the costs associated with the prediction is correlated with the processing costs. This view is compatible with the observation that the maintenance is the driving force for the parser to release the wh-filler from memory. For example, Gibson (1998) suggests that when the first NP within a clause is ambiguous between the dative case marker or the accusative case marker, the parser picks the accusative case because it requires less syntactic head to be integrated (Bader, Bayer, Hopf, & Meng, 1996; Gibson, 1998:59)

(7) a. Accusative ambiguous:

Dirigenten, die ein schwieriges Werk einstudiert haben, kann ein Kritiker ruhig umjubeln.

conductors who a difficult work rehearsed have can a critic safely cheer

'A critic can safely cheer conductors who have rehearsed a difficult work.'

b. Dative ambiguous:

Dirigenten, die ein schwieriges Werk einstudiert haben, kann ein Kritiker ruhig applaudieren.

conductors who a difficult work rehearsed have can a critic safely applaud

'A critic can safely applaud conductors who have rehearsed a difficult

---

[3] We assume that the storage and the maintenance are the same process.

work.'

c. Dative unambiguous:

Musikern, die ein schwieriges Werk einstudiert haben, kann ein Kritiker ruhig applaudieren.

musicians who a difficult work rehearsed have can a critic safely applaud

'A critic can safely applaud musicians who have rehearsed a difficult work.'

Specifically, when the initial NP is a dative, it predicts a verb which requires a subject and a direct object, marked with the nominative case marker and the accusative case marker, respectively. Choosing the dative indicates keeping the subject marked with the nominative case in memory until the prediction is satisfied (until another NP marked with the nominative case is encountered). This means that more syntactic heads need to be kept in memory, inducing memory costs. Therefore, the parser prefers to pick an accusative case marker when the initial NP is ambiguous in terms of the case marker.

Resolving Wh-Filler-Gap Dependencies (WhFGD) involves linking a wh-phrase to a verb, preposition, or gap. An example of a WhFGD construction is (8).

(8) Which mistake in the program will be disastrous for the company?

In (8), neither the interpretation nor the grammatical status of the wh-phrase *which mistake in the program* is determined solely by the wh-phrase itself. The wh-phrase, which is the subject of the *disastrous*, is interpreted as the theme argument of the predicate *disastrous*. In general, the grammatical status and the interpretation of a wh-phrase are determined in relation to other elements, such as the verb or preposition, or the gap, a controlling element. The dependent element is often referred to as a filler (e.g., the wh-phrase in *which mistake in the program*), and the controlling element which hosts the grammatically mandatory yet hidden argument is referred to as a gap.[4]

One of the important properties of long-distance dependencies is that they can span across a large number of words or clauses. In online WhFGD resolution, the parser needs to link the wh-filler to the gap in order to achieve the interpretation of the WhFGD sentence; for a wh-phrase to be interpreted, the wh-phrase needs to be linked to the gap. In other words, to resolve WhFGD, the parser needs to 'recover' the information of the filler after encountering the gap, in order to achieve the right interpretation of the sentence (Bever & McElree, 1988; Crain & Fodor, 1985; Fodor, 1978; Nicol et al., 1994). This implies that in order to resolve a WhFGD online, the parser needs to perform two processes. One is the storage or maintenance of a wh-filler (Gibson, 1998; Gibson & Warren, 2004; Wagers & Phillips, 2014; Wanner & Maratsos, 1978; Warren & Gibson, 2002), and the other is the retrieval or reactivation of the wh-filler (Lewis & Vasishth, 2005; McElree, 2001, 2006; McElree & Dosher, 1989; Nicol & Swinney, 1989; Oberauer & Kliegl, 2006; Van Dyke,

---

[4] Note that we do not commit to a specific analysis of WhFGD constructions. Specifically, we are agnostic about whether it involves a phonetically empty gap or not. We customarily call the controlling element as *gap*, but our conclusions do not necessarily require a gap-based analysis.

2007; Van Dyke & McElree, 2006).

2.1.2. Maintenance & Retrieval

Let us look at the maintenance and retrieval components in more detail. First, it is possible that the wh-filler is maintained in memory until the wh-filler is assigned a thematic role from the verb (Gibson, 1998; Gibson & Warren, 2004; Wagers & Phillips, 2014). Due to its morphological properties (i.e., *wh*-morphology, e.g., *which*) the parser can immediately recognize that the wh-filler is an element that will be linked to the gap somewhere downstream, or otherwise it cannot be interpreted. Note this is not true for other non-wh-NPs, like *the mistake*: a definite determiner *the* does not signal movement. The gap is not guaranteed to be adjacent to the wh-phrase, as it can appear in the subject position, the direct object position, the indirect object position, or the object position of a preposition. As such, the wh-phrase itself does not signal where the gap should be located. Thus, the parser needs to maintain the wh-filler in memory until the gap is identified and the wh-filler is successfully linked to the gap.

Numerous studies have shown that upon encountering the filler, the parser actively posits a gap in advance of confirming evidence. This is known as active dependency formation (Aoshima et al., 2004; Crain & Fodor, 1985; Frazier & Flores D'Arcais, 1989; Lee, 2004; Omaki et al., 2015; Phillips, 2006; Pickering & Barry, 1991; Stowe, 1986; Traxler & Pickering, 1996). Active dependency formation can be understood as a consequence of the parser's maintenance of a wh-filler in memory. That is, while a wh-filler must be linked to a gap, the distance between the filler and the gap is potentially long. This means that the parser needs to maintain the wh-filler in memory for a potentially long distance until the gap is encountered and the wh-filler can be linked to the gap. If the wh-filler is maintained in memory, it would be costly for long dependencies, which in turn would lead the parser to resolve the dependency as quickly as possible (Gibson, 1998; Gibson & Warren, 2004).

Evidence for maintenance comes from studies showing larger processing costs when the head of the dependency is not resolved immediately due to many intervening words. For example, Chen, Gibson, & Wolf (2005) showed that the readers have difficulty in maintaining multiple unresolved dependencies when the right-hand element of the dependency has not been encountered yet, such that the reading times at the most deeply embedded NP (*New York City*) position were faster for the relative clause constructions than the sentential complement constructions.

(9) a. SC structure: The announcement that the baker from a small bakery in New York City received the award helped the business of the owner.

b. RC structure: The announcement which the baker from a small bakery in New York City received helped the business of the owner.

This is because in sentential complement constructions, readers need to store the wh-element in memory until the dependency is resolved, whereas the wh-element in the relative clause does not need to be stored. In WhFGD processing, the reading time of the verb (that hosts the gap) is faster when the parser can form short filler-gap dependencies successively, versus when the parser needs to hold the filler for a longer time (see Keine, 2015 for related observations). This claim is bolstered by the findings of Gibson and

Warren (2004), who observed that the reading times were slower when the number of words intervening between the wh-filler and the gap increased. When sentences involving wh-extraction are compared to those that do not involve wh-extraction, reading times of the words between the filler and gap increase for longer dependencies (see also Chen et al., 2005 and Stepanov & Stateva, 2015).

Once the gap is recognized, information associated with the wh-filler needs to be recovered or retrieved (McElree, 2006; McElree, Foraker, & Dyer, 2003; McElree & Dosher, 1989). This is necessary for the parser to check the case, thematic role and other morphological features of the filler and to achieve its proper interpretation (Bever & McElree, 1988; Crain & Fodor, 1985; Fodor, 1978; McElree & Bever, 1989; Nicol et al., 1994; Nicol & Swinney, 1989). Fillers may contain different kinds of information, including morphological features, syntactic category, and lexical-semantic content. Some of these properties may be subject to memory decay (King & Just, 1991; Wagers & Phillips, 2014; Wanner & Maratsos, 1978). Different information could be maintained during the resolution of the dependency, or could decay and then be retrieved when the gap is recognized.[5]

---

[5] As an anonymous reviewer in *Language, Cognition, and Neuroscience* points out, some previous studies have suggested that decay is no longer a useful explanatory concept in the retrieval literature (see Berman, Jonides, & Lewis, 2009; Lewandowsky, Geiger, & Oberauer, 2008; McElree, 2006; Nairne, 2002;). Wagers and Phillips (2014) pointed out that not all the features of the elements that are retrieved at the head of the dependency or are fully reactivated at the verb position (e.g., semantic features of the wh-filler in Wagers and

Wagers and Phillips (2014) investigated which aspects of the filler are maintained, and which are susceptible to decay. They observed a filled-gap effect (e.g. Boland, Tanenhaus, Garnsey, & Carlson, 1995; Crain & Fodor 1985; Frazier & Clifton, 1989; Stowe, 1986; Tanenhaus, Boland, Garnsey, & Carlson, 1989; Wagers & Phillips, 2014) when the wh-filler is an NP and the potential gap site is also an NP, but not when the wh-filler is a PP and the potential gap site is an NP, regardless of whether the WhFGD spans a short or long distance, as illustrated in (10) (Wagers & Phillips, 2014; 1282); n.b. *FGE* stands for *filled gap effect*.

(10)

					Plausibility Mismatch. ↓		*FGE* ↓
a.	Wh-NP	...	SHORT DISTANCE	...	**V**	...	**NP**
b.	Wh-PP	...	SHORT DISTANCE	...	V	...	NP
							FGE ↓
c.	Wh-NP	...	LONG DISTANCE	...	V	...	**NP**

Phillips's (2014) study). Such findings can be accounted for by memory decay. Thus, for present purposes, we hypothesize that some of the information associated with the filler is subject to decay or interference. We assume that the success of retrieval is related to the amount of material intervening between the filler and the gap.

d. Wh-PP ... LONG DISTANCE ... V ... NP

This suggests that category information of the wh-filler is maintained throughout the dependency formation process. However, the semantic incongruity between the wh-filler and the verb (e.g. Boland et al., 1995; Traxler & Pickering, 1996) was not recognized. That is, the readers cannot detect the semantic incompatibility of the filler and the verb when the dependency spans a long distance, nor is the mismatch between the preposition attached to the wh-phrase and the verb recognized in long distance dependencies. This suggests that syntactic category information of the fillers is maintained during the online WhFGD formation process, but semantic content and lexical information are released from maintenance. In sum, the implication is that resolving filler gap dependencies involves both maintenance and retrieval, and the information that is retrieved at the verb position reflects what information of the filler is maintained and what information of the filler is released from maintenance.

Note that Chow & Zhou (2019) recently suggested that the lack of a plausibility effect is not because the content of the wh-filler is released from maintenance, but because of the lack of statistical power in earlier studies. They conducted an eye-tracking experiment with high statistical power. Like Wagers and Phillips (2014), they found that readers actively insert a gap regardless of dependency length whenever one is grammatically possible, suggesting an active gap filling effect. In addition, they found a plausibility effect in regression path duration as well as total reading times for all dependency lengths. Their findings therefore provide evidence for the maintenance of semantic features. Furthermore, they found a weaker plausibility mismatch effect after the critical region for long dependencies, relative to short dependencies. Therefore, it is possible that, contrary to Wagers and Phillips (2014), thematic information can be maintained in memory. However, distance still impacts the retrieval of thematic information as the observed plausibility effects show, which suggests that memory decay may be in effect.

In the current studies, we investigate this claim, asking what sort of information from the wh-filler can be maintained: just category information, or something more detailed? Through a series of studies on online WhFGD formation, we show that, like in Chow & Zhou (2019), the wh-filler needs to be maintained in memory throughout the processing of WhFGD sentences, but if the wh-filler is released from maintenance and retrieved later, the relative strength of the filler, and thus the degree of success of its retrieval, is reduced.

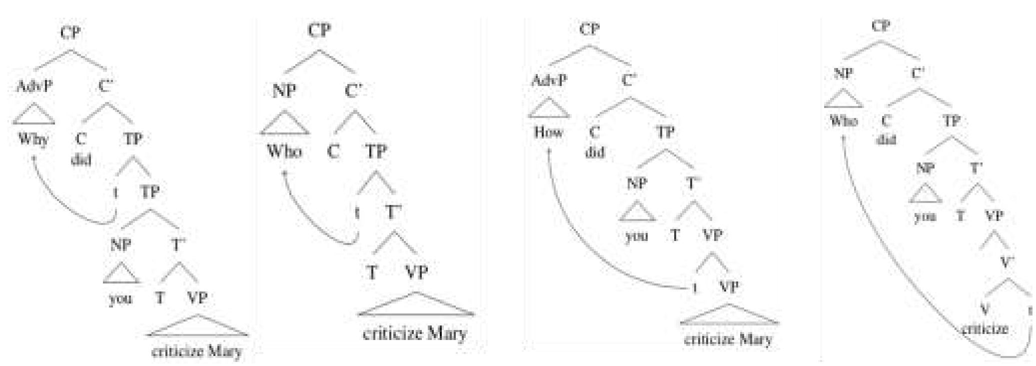
We also posit an additional question: what motivates the maintenance of an element? In the case of WhFGD, wh-fillers involve distinctive wh-morphology. In a language like English, which is a wh-movement language, a phrase bearing wh-morphology provides strong evidence for the presence of WhFGD, i.e., if there is a wh-phrase, there must be a gap somewhere in the sentence (Chomsky, 1977). Therefore, it is plausible that wh-morphology signals the presence of a filler-gap dependency and thus leads the parser to maintain the wh-phrase. On the other hand, when a phrase does not bear wh-morphology, it is unclear whether the phrase is part of a filler-gap dependency. Assuming that maintaining an element in memory is costly (Gibson, 1998; Wanner & Maratsos, 1978), it is plausible that the parser does not maintain non-wh-phrases in memory in the same way as wh-phrases. We investigate these points by examining in detail the processing of coordinated structures involving WhFGDs.

2.2. *Grammatical Properties and the Processing of Wh-Phrases (WhPs)*

2.2.1. WhPs

Recent syntactic studies on wh-constructions have revealed that each wh-phrase has different controlling elements and thus form a dependency with different elements (Chomsky, 1986; Huang, 1982; Lasnik & Saito, 1984; Stepanov & Tsai, 2008). Syntactically, WhPs differ in their movement profiles and in which controlling element they form a dependency with. *Why* is known to be base-generated in the CP position, or to undergo a short movement from TP to CP (Ko, 2005; Yoshida et al., 2015) and form a dependency with TP. *Who*, functioning as the subject of the sentence (subject *who*: who_{subj}) moves from TP to CP, and forms a dependency with TP. *How*, a modifier of VP, and *who* (object *who*: who_{obj}), an argument of V, both undergo movement from within VP and move to CP. Although these WhPs appear in the same linear position, the levels with which each wh-phrase are associated are different.

(11) a. b. c. d.

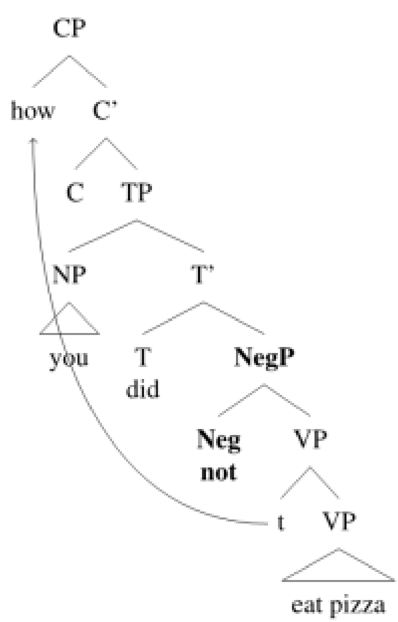


Evidence for *why* forming a dependency with TP comes from its insensitivity to inner-islands like negative-islands (Ross, 1984). Ross (1984) observes that when a wh-gap dependency spans across a sentential negation, *not*, the acceptability of such dependency is degraded. This negative island effect leads to the contrast between (12a) and (12b). It has been known that *how*, which is understood as a modifier for VP, needs to undergo movement from the VP-domain across negation to higher in the CP position, and is therefore sensitive to negative islands. On the other hand, *why*-questions are insensitive to negative islands, which suggests that *why* is base-generated in CP or undergoes a short movement from TP to CP, and thus does not move across negation.

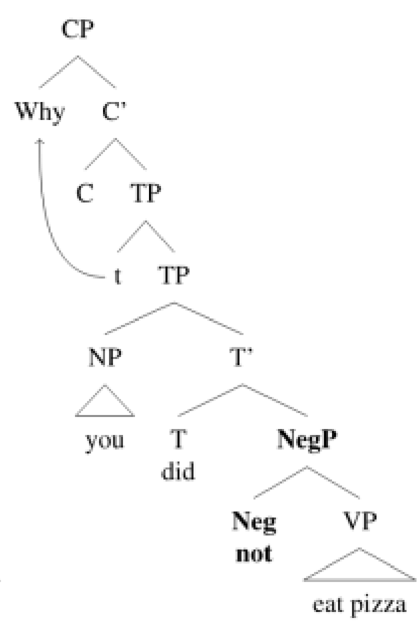
(12) a. I don't know why you did not eat pizza.

b. *I don't know how you did not eat pizza?

c.



d.



Who$_{subj}$ forms a dependency with TP but *how* and *who*$_{obj}$ form dependencies with VP or V. This can be shown by the difference in acceptability in the following examples in (13). It has been known that the wh-gap dependency that spans across another wh-element gives rise to the acceptability degradation (the wh-island effect: Chomsky, 1977, Sprouse, Wagers, & Phillips, 2012; Kush, Lohndal, & Sprouse, 2018, for detailed experimental investigations). The acceptability contrast in (13) suggests that *how* and *who_{obj}* both move from the subordinate clause to the matrix CP position, thus the dependencies span across a whP, *why*, which gives rise to the acceptability degradation. On the other hand, *who_{subj}* moves from the matrix TP to the matrix CP, and thus the wh-gap dependency does not span across another whP, and (13a) is more acceptable than (13b-c).

(13) a. Who t wonders **why** John criticized Mary?

b. *How did you wonder **why** John t criticized Mary?
(intended answer: *harshly*)

c. *Who did you wonder **why** John criticized t?
(intended answer: *Mary*)

Furthermore, when the subject position (which is in TP) is occupied by an NP, an example of *who_{subj}* no longer becomes acceptable, suggesting that in such an example, the *who_{subj}* does not have an element to form a dependency with, i.e., that *who_{subj}* is indeed linked to TP.

(14) *Who John criticized Mary?

In the same token, when the adverb position is occupied by a manner adverb, a phrase using *how* is not acceptable (when *how* is interpreted as a manner adverb); when the object position is occupied by an NP, an example with who_{obj} is not acceptable. These observations support the analysis that *how* is linked to VP and who_{obj} is linked to V.

(15) a. *How did John harshly criticize Mary?

b. *Who did John criticize Mary?

2.2.2. Time Course

With this in mind, let us walk through a time-course of the processing of different WhPs.

(16) a. Who did Anna criticize?

b. How did Anna criticize?

c. Why did Anna criticize?

d. Who criticize Anna?

The dependency lengths for these WhPs are different, and thus based on the dependency lengths, different processing costs are predicted. One of the ways to empirically test how different fillers are associated with different degrees of processing cost is to investigate maintenance cost effects (Gibson, 1998; Gibson & Warren, 2004). In addition, the costs associated with the maintenance of those fillers should be released once the WhFGD formation is completed, no longer impacting memory resources. Based on this logic, we can observe differences in the complexity measure depending on the length of the dependency. In order to investigate the complexity effect between WhPs and the verb, we examine how different dependency types interact with a complex domain intervening between the WhP and the licensor.

Let us observe what happens to different kinds of WhPs (*who*$_{obj}$, *how*, and *why*) when a complex domain like (17a) is added in the context of the WhFGD construction as in (17b-d).

(17) a. [$_{NP}$ the babysitter [$_{CP}$ that the children loved]] handed the toys to someone.

b. The father considered <u>who</u> the babysitter that the children loved handed the toys to.

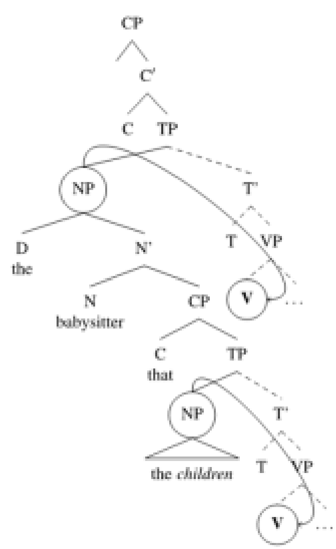
c. The father considered <u>how/that/why</u> the babysitter that the children loved handed the toys to the grandma.

d. <u>Who</u> considered that the naive babysitter that the spunky children loved handed the toys to the grandma.

(18a) is an example of an object-gapped relative clause that is modifying the subject NP. This type of center-embedded relative clause configurations is known to create a processing cost. As the diagram in (18b) illustrates, in a configuration like (18a), there are two open dependencies at the most deeply embedded noun *children*: the dependency between [$_{NP}$ *the babysitter*] and the matrix verb, and [$_{NP}$ *the children*] and the embedded verb. Under theories of processing complexity in which the number of open dependencies is a predictor of processing complexity (the processing complexity is associated with the storage cost: Chen et al., 2005; Gibson 1998, 2000; Grodner & Gibson, 2005; Stepanov & Stateva, 2015), processing complexity is predicted to be highest at that point, the word *children*.

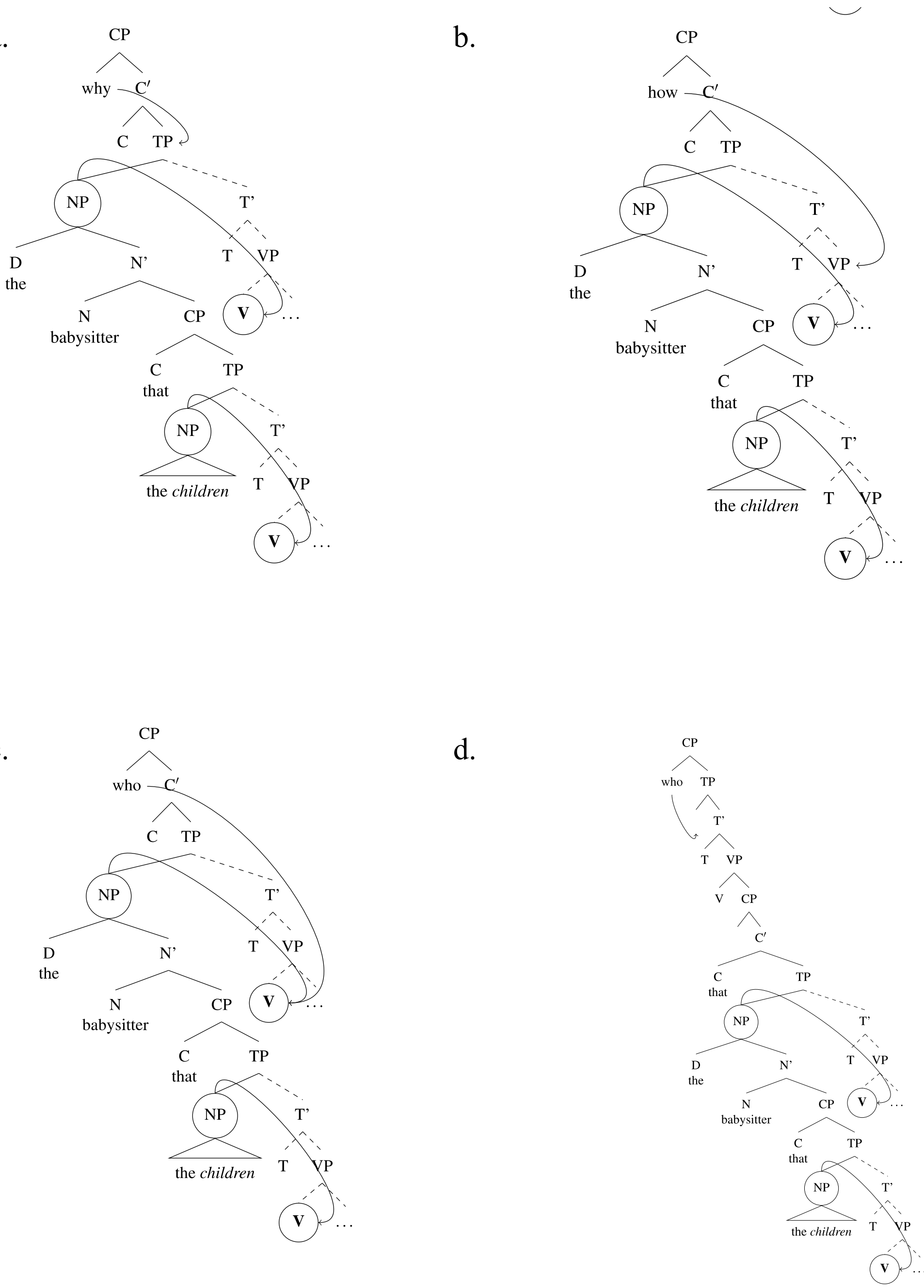
(18) a. [$_{NP}$ The babysitter [$_{CP}$ that the children loved]] handed a toy to the baby.

b.



Different degrees of complexity can be observed at the most deeply embedded noun position when the complex NP is embedded within different WhFGD constructions. The structures of each WhFGD construction with the complex NP embedded in the middle are summarized in (19).

(19) a. b. c. d.



As we can see in (19a), *why* forms a dependency with TP, and, at the point of *children*, the *why*-TP dependency has been already completed and thus there are only two open dependencies. In the WhFGD construction involving *how* as in (19b), which forms a dependency with the VP, there are three open dependencies at *children* (the dependency between *how* and VP, [$_{NP}$ *the babysitter*] and V, and [$_{NP}$ *the children*] and V). In the construction involving *who*$_{obj}$, which is linked to the verb as in (19c), there are three open dependencies at the point of *children* (*who*$_{obj}$ and V, [$_{NP}$ *the babysitter*] and V, and [$_{NP}$ *the children*] and V). Finally, in the construction that involves *who*$_{subj}$, which is linked to TP, as illustrated in (19d), there are two open dependencies at the point of *children*. If the number of open dependencies predicts processing complexity, then we expect that processing complexity at the point of *children* would be higher in the *how* and *who*$_{obj}$ constructions than in the *why* and *who*$_{subj}$ constructions. If processing complexity is linked to a reading time slowdown in online processing (Gibson, 1998; Gibson & Warren, 2004), then we expect that *children* in the *how* and *who*$_{obj}$ constructions should be read significantly slower than in the *why* and *who*$_{subj}$ constructions.

2.3. Motivation for the Experiments (Experiment 1 & 2)

We conducted four acceptability rating experiments (experiments 1a-1d) and two self-paced reading experiments (experiment 2a & 2b) to understand how different WhPs with different dependency lengths contribute to different storage costs. The first acceptability rating experiment tested whether the experimental sentences used in experiment 2a were plausible without errors. The second acceptability rating experiment tested whether the experimental sentences in experiment 2b were plausible without errors. The third acceptability rating experiment was designed to test whether *why* behaves differently from other WhPs such that *why* is not linked to the verb, by examining the sensitivity of *why* to negative islands. The fourth experiment tested whether there are two different *why*s with different structural heights in English (Chapman & Kučerová, 2016). The online experiments were designed to test whether different kinds of WhPs are released from memory once the WhFGD is formed.

2.4. Different Kinds of WhPs: An Acceptability Rating Experiment 1a

The goal of the experiment 1a is to examine how the sentences with who_{obj}, *how*, *that* and *why* were rated in offline experiments.

2.4.1. Participants, Materials and Design

For this experiment, 45 native English speakers from Northwestern University with no history of language disorders participated and gave informed consent. In exchange for their participation, the participants were granted 1 credit for introductory linguistic classes taught at Northwestern.

Critical items consisted of 24 sentence sets in the form of a 1x4 design, in which different kinds of WhPs (who_{obj}, *how*, *that*, *why*) were manipulated as independent factors. A sample set of stimuli is summarized in Table 1. To ensure that participants did not encounter the same types of target items consecutively, items were distributed in a pseudo-randomized manner. In addition to the current experimental items, there were 32 filler sentences that involved irrelevant manipulations to the current ones. The experiment took around 30 minutes to complete.

Table 1. Sample stimuli for experiment 1a.

Factors WhPs	Examples
who$_{obj}$	The father considered who the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to.
how	The father considered how the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma.
that	The father considered that the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma.
why	The father considered why the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma.

2.4.2. Procedure

Stimuli were presented on a desktop PC using Linger software (Rohde, 2003). For each stimulus, participants read the sentence on a screen, and were directed to rate the sentences from 1 to 7 with regard to their naturalness (1: totally unacceptable, 7: totally acceptable). To familiarize participants with the rating process, five practice items were presented before the actual experimental items. Participants were instructed that they did not need to read aloud and that there are not necessarily right answers to each question.

2.4.3. Analysis

Data were analyzed using linear mixed effect regression models performed with the *lme4* package in R version 3.2.3 (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Bates, Maechler, Bolker, & Walker, 2014; Jaeger, 2008). Each model included helmert coding where we compared (i) *who*$_{obj}$/*how*/ *why* with *that* as a baseline, (ii) *who*$_{obj}$/ *how* with *why*, and (iii) *who*$_{obj}$ and *how*. All models contained the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013), which involved random intercepts for participants and items and random slopes for fixed effects assuming the model converged. In cases where the model failed to converge, the random effects with the least variance were taken out step-wise.

2.4.4. Results & Discussion

Mean acceptability scores are shown in Table 2, and in Figure 1, and mixed effect model outputs are shown in Table 3. The results only revealed a significant effect between *who*$_{obj}$ and *how*.

Table 2. Mean acceptability ratings from experiment 1a.

WhPs	Mean	SE
who$_{obj}$	3.64	0.10
how	4.32	0.10
that	4.08	0.10
why	4.01	0.11

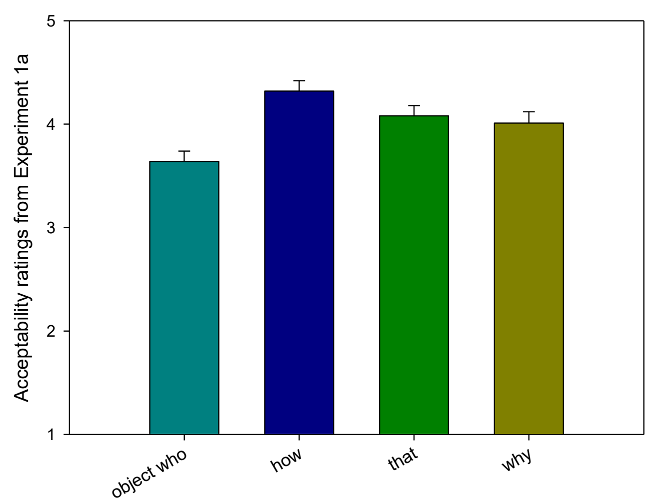


Figure 1. Mean acceptability ratings from experiment 1a.

Table 3. Summary of mixed effect model outputs in experiment 1a.

WhPs	Estimate (SE)	t-value
(Intercept)	4.05 (0.17)	23.75
who$_{obj}$/*how*/*why* vs. that	0.07 (0.19)	0.77
who$_{obj}$/*how* vs. *why*	0.02 (0.07)	0.35
who$_{obj}$ vs. *how*	-0.34 (0.08)	-4.20 ***

2.5. Different Kinds of WhPs: An Acceptability Rating Experiment 1b

The goal of experiment 1b is to examine how the sentences containing *who*$_{obj}$, *who*$_{subj}$, *that*, and *why* are rated in offline experiments.

2.5.1. Participants, Materials and Design

For this experiment, 24 native English speakers from Northwestern University with no history of language disorders participated and gave informed consent. In exchange for their participation, the participants were granted 1 credit for introductory linguistic classes taught at Northwestern.

Critical items consisted of 32 sentence sets arranged in a 1x4 design, in which different kinds of WhPs (*who*$_{obj}$, *who*$_{subj}$, *that*, *why*) were manipulated as independent factors. A sample set of stimuli is summarized in Table 4. To ensure that participants did not encounter the same types of target items consecutively, items were distributed in a pseudo-randomized manner. In addition to the current experimental items, there were 32 filler sentences that involved irrelevant manipulations to the current ones. The experiment lasted around 30 minutes for each participant.

Table 4. Sample stimuli for experiment 1b.

Factors	
WhPs	Examples
who_{obj}	The father considered who the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to.
who_{subj}	Who considered how the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma?
that	The father considered that the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma.
why	The father considered why the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma.

2.5.2. Procedure

The same procedure was employed as in Experiment 1a.

2.5.3. Analysis

Each model included helmert coding where we compared (i) who_{obj}/ who_{subj}/*why* with *that* as a baseline, (ii) who_{obj} with who_{subj}/*why*, and (iii) who_{subj} and *why*.

2.5.4. Results & Discussion

Mean acceptability scores are shown in Table 5 and in Figure 2, and mixed effect model outputs are shown in Table 6. The results revealed a significant difference between who_{obj} /who_{subj}/*why* vs. *that*. There was also a significant difference between who_{obj} vs. who_{subj}/*why*. There was no significant difference between who_{subj} and *why*.

Table 5. Mean acceptability ratings from experiment 1b.

WhPs	Mean	SE
who$_{obj}$	3.32	0.14
who$_{subj}$	3.75	0.13
that	3.99	0.13
why	3.77	0.14

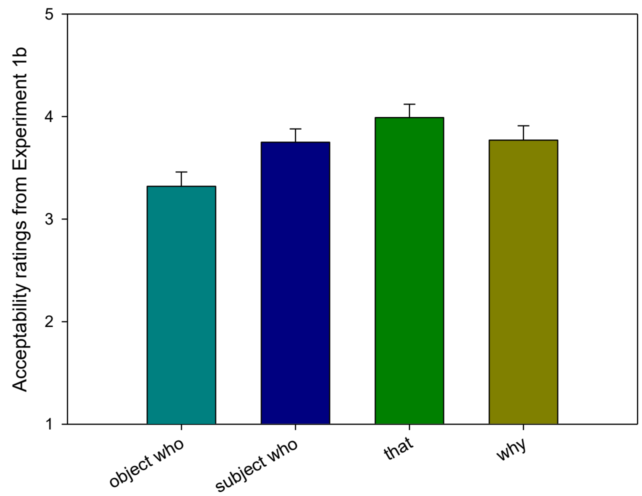


Figure 2. Mean acceptability ratings from experiment 1b.

Table 6. Summary of mixed effect model outputs in experiment 1b.

WhPs	Estimate (SE)	t-value
(Intercept)	3.71 (0.23)	15.91
who$_{obj}$/who$_{subj}$/why vs. that	0.28 (0.11)	2.59 *
who$_{obj}$ vs. who$_{subj}$/why	-0.30 (0.10)	-3.02**
who$_{subj}$ vs. why	-0.02 (0.08)	-0.24

2.6. Different Kinds of WhPs: An Acceptability Rating Experiment 1c

The goal of the experiment 1c is to further test whether *how* behaves similarly with *who$_{obj}$* in its grammatical properties in comparison to *why*. This was tested by the acceptability judgment of Question-Answer pairs for different WhPs in negative island contexts, as shown in Table 7. The answers for *who$_{obj}$* were the direct objects of the verb, the answers for *how* were adverbs, and the answers for *why* started with a Because-clause. The prediction is that the mean ratings of Question-Answer pairs for *how* would be lower than those of *why*, because *how* is located in the VP position rather than in the TP position, and because *how* is known to be sensitive to negative islands. This experiment examines whether *why* is licensed by TP, and whether it is generated in a higher position than other WhPs. If *why* is not sensitive to inner islands, like negative islands, this would indicate that the licensor for *why* is above the TP position and thus does not interfere with negative clauses. This is because *why* is base-generated in CP or undergoes a short movement from TP to CP, which does not move across negation. These properties are not predicted if *why* moves from the VP area like other wh-phrases do. On the other hand, licensors for who_{obj} and *how* are linked to VP areas and therefore interfere with negative clauses. Thus, we predict that negative islands with *why* will have significantly higher acceptability ratings than who_{obj} and *how*.

2.6.1. Participants, Materials and Design

For this experiment, 40 Northwestern University students who were native English speakers with no history of language disorders gave informed consent and participated. In exchange for their participation, the participants were granted 1 credit necessary for introductory linguistic classes at Northwestern.

Critical items consisted of 24 sentence sets in the form of a 1x4 design, in which different kinds of WhPs (who_{obj}, *how*, *that*, *why*) were manipulated as independent factors. In addition to the current experimental items, there were 32 filler sentences that involved irrelevant manipulations to the current ones. A sample set of stimuli is summarized in Table 7.

Table 7. Sample stimuli for experiment 1c.

Factors	
WhPs	Examples
who_{obj}	A: Who didn't Mary criticize? B: Tommy.
how	A: How didn't Mary criticize John? B: Harshly.
that	A: Didn't Mary criticize John? B: Yes, she did.
why	A: Why didn't Mary criticize John? B: Because his answer was correct.

2.6.2. Procedure

The same procedure was used as in Experiment 1a.

2.6.3. Results & Discussion

Mean acceptability scores are shown in Table 8 and in Figure 3, and mixed effect model outputs are shown in Table 9. Overall, the results revealed a significant difference between who_{obj}/*how*/*why* vs. *that*. There was also a significant difference between who_{obj}/*how* vs. *why*. Finally, the results revealed differences between who_{obj} vs. *how*. This suggests that *why* is insensitive to negative islands as it does not move.

Table 8. Mean acceptability ratings from experiment 1c.

WhPs	Mean	SE
who_{obj}	6.10	0.08
how	2.99	0.12
that	6.25	0.09
why	6.46	0.07

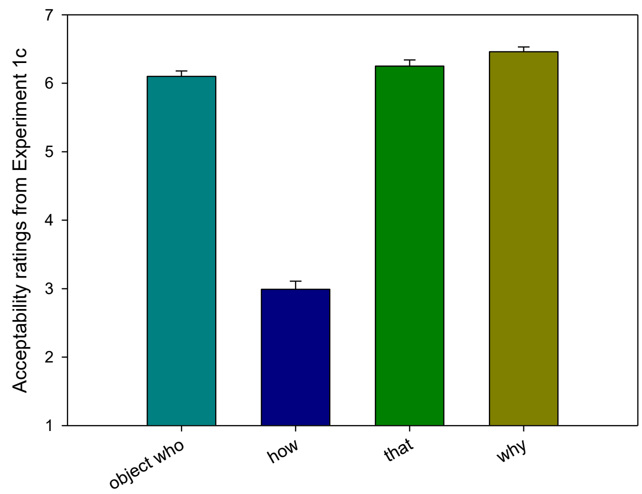


Figure 3. Mean acceptability ratings from experiment 1c.

Table 9. Summary of mixed effect model outputs in experiment 1c.

WhPs	Estimate (SE)	t-value
(Intercept)	5.46(0.11)	51.36
who$_{obj}$/how/why vs. that	0.74(0.14)	5.14 ***
who$_{obj}$/how vs. why	1.27 (0.12)	10.55 ***
who$_{obj}$ vs. how	1.58(0.13)	12.40 ***

2.7. Different Kinds of WhPs: An Acceptability Rating Experiment 1d

The goal of the experiment 1d is to empirically test whether there are two different *why*s that have different structural positions in English. This was tested by the acceptability ratings of Question-Answer pairs for different WhPs in negative island contexts, similar to experiment 1c. However, the answers for *why* either started with Because-clause (we call this as *reason why*: *why_$_R$*) or In order to-clause (we call this as *purpose why: why_$_P$*). The rationale behind this experiment was the following: Chapman & Kučerová (2016) propose that there are two different *why*s in English that correspond to different heights in structures: *why_$_R$* located at the Spec_CP and *why_$_P$* in the VP area, based on the evidence that *why_P* is sensitive to negative islands because it crosses the NegP, but *why_R* is not sensitive to negative islands. Thus, the prediction is if there are two different *why*s, one that is located at the Spec_CP position, and the other in VP area, then we expect Q/A pairs that elicit *why_R* (Because-clause) to yield higher acceptability ratings than *why_P* (In order to-clause). On the other hand, if there are no two different *why*s, and both of them are located in the Spec_CP position, we expect no differences between the mean acceptability ratings of *why_R* and *why_P*.

2.7.1. Participants, Materials and Design

For this experiment, 43 native English speakers from Northwestern University with no history of language disorders participated and gave informed consent. In exchange for their participation, the participants were granted 1 credit for introductory linguistic classes taught at Northwestern.

Critical items consisted of 32 sentence sets in the form of a 1x4 design, in which different kinds of WhPs (*how, that*, *why_R* , *why_P*) were manipulated as independent factors. In addition to the current experimental items, there were 32 filler sentences that involved irrelevant manipulations to the current ones. A sample set of stimuli is summarized in Table 10.

Table 10. Sample stimuli for experiment 1d.

Factors	

WhPs	Examples
how	A: How didn't Mary criticize John? B: Harshly.
that	A: Didn't Mary criticize John? B: Yes, she did.
reason why ($why\_{}_{R}$)	A: Why didn't Mary criticize John? B: Because his answer was correct.
purpose why ($why\_{}_{P}$)	A: Why didn't Mary criticize John? B: In order to make John angry.

2.7.2. Procedure

The same procedure was used as in Experiment 1a.

2.7.3. Results & Discussion

Mean acceptability scores are shown in Table 11 and in Figure 4, and mixed effect model outputs are shown in Table 12. Overall, the results revealed a significant difference between *how*/$why\_{}_{R}$ /$why\_{}_{P}$ vs. *that*. It also revealed a significant difference between *how* vs. $why\_{}_{R}$ /$why\_{}_{P}$. Importantly, there was also a significant difference between $why\_{}_{R}$ and $why\_{}_{P}$. This suggests that $why\_{}_{R}$ is insensitive to negative islands as it does not move, but $why\_{}_{P}$ is sensitive to them.

Table 11. Mean acceptability ratings from experiment 1d.

WhPs	Mean	SE
how	3.01	0.07
that	6.50	0.05
$why\_{}_{R}$	6.61	0.04
$why\_{}_{P}$	5.74	0.07

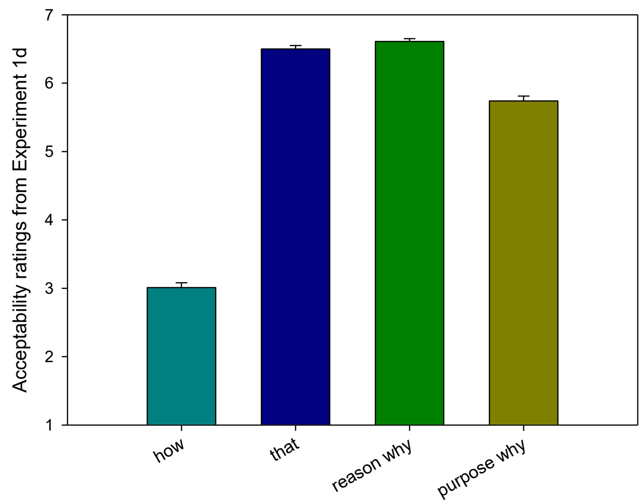


Figure 4. Mean acceptability ratings from experiment 1d.

Table 12. Summary of mixed effect model outputs in experiment 1d.

WhPs	Estimate (SE)	t-value
(Intercept)	5.47 (0.10)	56.47
how/why_R/why_P vs. that	1.04 (0.14)	7.59 ***
how vs. why_R/why_P	-2.12 (0.15)	-14.11 ***
why_R vs. *why_P*	-0.43 (0.08)	-5.14 ***

2.7.4. Summary of Experiments 1a, b, c, & d

The purpose of the first two experiments was to examine the overall naturalness of the sentences we would like to use in the online experiments. Although the sentences were complex, the average ratings for each wh-phrase were relatively high for both experiment 1a and 1b. The Q/A dialogue was employed to understand whether different WhPs are sensitive to negative islands. We showed that *why* is insensitive to negative islands in contrast to *who*, and that there was a significant difference between *why_R* and *why_P*. The insensitivity to negative islands for *why* further suggests that *why* is located higher in the structure than other wh-phrases. Furthermore, the significant difference between *why*$_R$ and *why*$_P$ suggests that these two different *why*s can be located in different positions.

2.8. Maintenance of Different Kinds of WhPs: A Self-Paced Reading Experiment 2a

This study investigates the role of maintenance in holding different kinds of WhPs by adding a complex domain between different WhPs and the verb. Assuming that the processing complexity is linked to the slower reading times (Chen et al., 2005; Gibson 1998, 2000; Gibson & Warren, 2004; Grodner & Gibson, 2005), we predicted that the reading times for *who*$_{obj}$ and *how* will be longest at the embedded NP position whereas *why* will not increase complexity at the embedded NP, as *why* is already released at the point of TP, and thus will not increase reading times. That is, given that the filler is released from memory once the dependency is formed, we predict no differences in reading times for *how* and *who*$_{obj}$, but *why* should have the fastest reading times at the preverbal position (embedded NP) as *why* is released from memory upon encountering TP.

2.8.1. Participants, Materials and Design

For this experiment, 64 Northwestern University students who were native English speakers with no history of language disorders gave informed consent and participated. In exchange for their participation, the participants were granted 1 credit necessary for introductory linguistic classes at Northwestern.

Critical items consisted of 24 sentence sets in the form of a 1x4 within-subjects design, in which different kinds of WhPs (*who*$_{obj}$, *how*, *that*, *why*) were manipulated as independent factors. A sample set of stimuli is summarized in Table 13. To ensure that participants did not encounter the same types of target items consecutively, items were distributed in a pseudo-randomized manner. In addition to the current experimental items, there were 74 filler sentences that involved irrelevant manipulations to the current ones. The experiment took around 30 minutes to complete.

Table 13. Sample stimuli for experiment 2a

Factors	
WhPs	Examples
who$_{obj}$	The father considered who the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to.
how	The father considered how the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma.
that	The father considered that the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma.
why	The father considered why the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma.

2.8.2. Procedure

Stimuli were presented on a desktop PC using Linger software (Rohde, 2003) where the experiment followed a self-paced word-by-word moving window paradigm (Just, Carpenter, & Woolley, 1982). Each experimental trial started with dashes masking the words in the sentence, and the participant pressed the button to move forward. Participants were instructed to read the sentences as they would normally read, and to answer comprehension questions after reading each sentence. The yes/no comprehension question asked participants to press F (yes) or J (no) keys. An example comprehension question is *Was the word toys mentioned* *in the story?*. After receiving feedback on their response, participants pressed the space bar to proceed to the next item. Six practice items were presented at the beginning of the experiment. The experiment took each participant about 30-45 minutes to complete.

2.8.3. Analysis

Data were analyzed using linear mixed effect regression models performed with the *lme4* package in R version 3.2.3 (Baayen, 2008; Baayen et al., 2008; Bates et al., 2014; Jaeger, 2008). Each model included helmert coding where we compared (i) who_{obj}/*how*/*why* with *that* as a baseline, (ii) who_{obj}/*how* with *why*, and (iii) who_{obj} vs. *how*. All models contained the maximal random effects structure that fit the data (Barr et al., 2013), including random intercepts for participants and items and random slopes for fixed effects given the model successfully converged. In cases where the model did not converge, the random effects that revealed the least variance were taken out in a step-wise manner. Participants whose accuracy was lower than 68% were excluded. Reading times were log-transformed with an aim to minimize non-normality (Box & Cox 1964; Vasishth, Chen, Li, Guo, & Paterson, 2013). The critical regions were the most deeply embedded NP (the embedded NP critical region) and one word following the critical region (the embedded NP spillover region). The second critical regions were the matrix verb, and one word following the matrix verb (the matrix verb spillover region).

2.8.4. Results & Discussion

Region-by-region reading times are presented in Figure 5, the graph at the critical region (the most deeply embedded NP) in Figure 6, and the graph at the matrix verb spillover region in Figure 7. The mixed effect model outputs are presented in Table 14. Mean accuracy for critical trial comprehension questions was 77.0%.

Table 14. Summary of mixed effect model outputs for experiment 2a.

Region		Estimate (SE)	t-value
Embedded NP Critical Region	Intercept	5.82 (0.03)	189.16
	who$_{obj}$/how/why vs. that	-0.01 (0.02)	-0.32
	who$_{obj}$/how vs. why	-0.03 (0.01)	-2.16 *
	who$_{obj}$ vs. how	0.01(0.01)	0.38
Embedded NP Spillover Region	Intercept	5.98 (0.04)	167.80
	who$_{obj}$/how/why vs. that	-0.03 (0.02)	-1.80
	who$_{obj}$/how vs. why	-0.00 (0.03)	-0.07
	who$_{obj}$ vs. how	0.01 (0.02)	0.86
Matrix Verb Critical Region	Intercept	5.98 (0.04)	168.01
	who$_{obj}$/how/why vs. that	-0.03 (0.02)	-1.81
	who$_{obj}$ vs. how/why	0.01 (0.02)	0.76
	how vs. why	-0.01 (0.02)	-0.37
Matrix Verb Spillover Region	Intercept	5.77 (0.03)	221.05
	who$_{obj}$/how/why vs. that	-0.01 (0.01)	-0.63
	who$_{obj}$ vs. how/why	0.03 (0.01)	2.17 *
	how vs. why	0.01 (0.01)	1.17

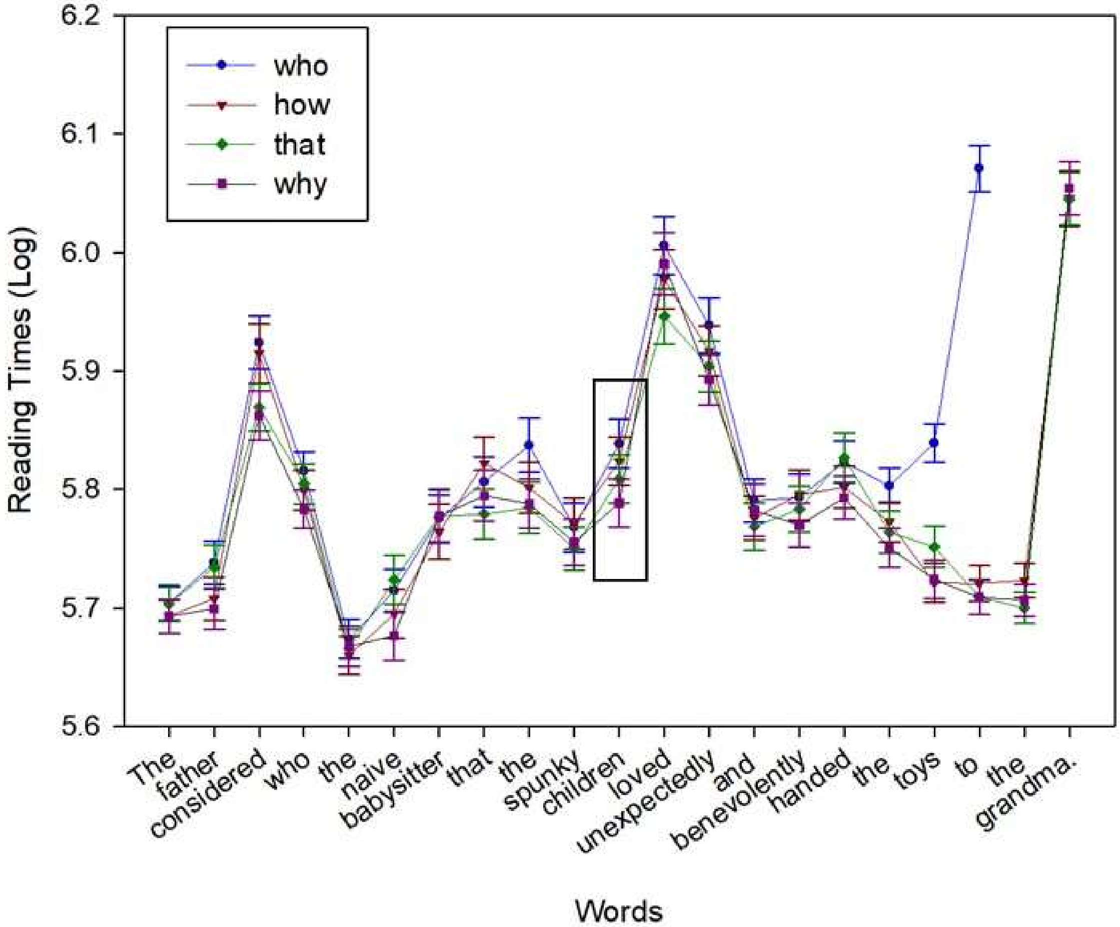


Figure 5. Region-by-region reading time means from experiment 2a.

The results revealed that at the embedded NP, children, *who*$_{obj}$ and *how* were read significantly slower than *why* while there was no significant difference between *who*$_{obj}$ and *how*. There were no significant differences between *who*$_{obj}$/*how*/*why* vs. *that*[6]. The region

---

6 We also took one region prior to the critical region as a predictor to check the covariance between the prior region and the critical region; the covariance did not affect the results that much (*who*$_{obj}$/*how* vs. *why*: Estimate: -0.03, SE= 0.01, t=-1.97).

after the embedded NP revealed no significant effect. It seems that *why* does not impact the reading time of the RC subject compared to other WhPs. The difference between *why* and *who*$_{obj}$/*how* shows that these WhPs are processed differently, in the direction we expected given the grammatical differences between them, and that readers maintain different WhPs in memory.

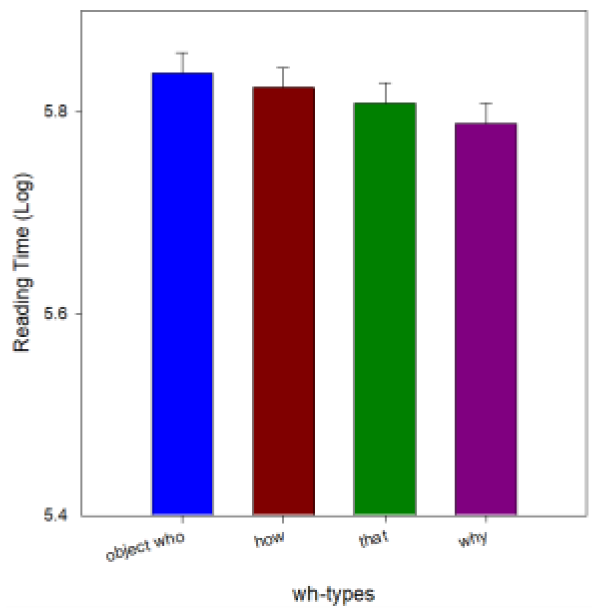


Figure 6. Reading times at the embedded NP region in experiment 2a.

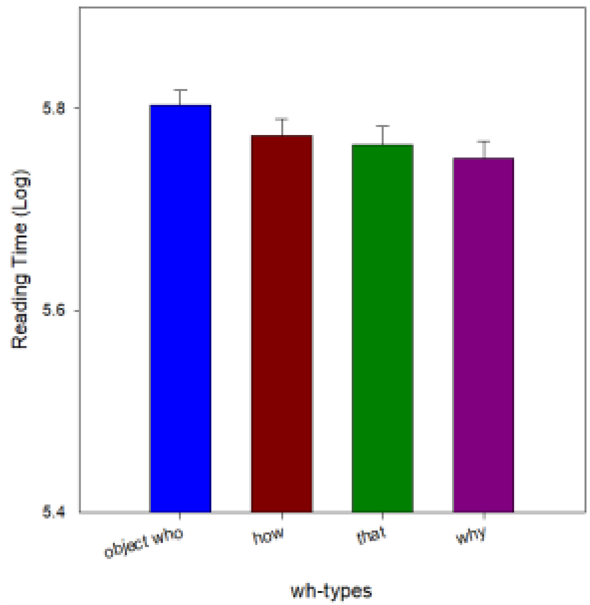


Figure 7. Reading times at the matrix verb spillover region in experiment 2a.

At the matrix verb spillover region, however, who_{obj} was read signficantly slower than *why* and *how*, but no differences were observed between *how* and *why*[7]. Given that who_{obj} is an argument of the verb, *handed*, and *how* is a modifier of the VP, the dependency length from who_{obj} to the verb, *handed*, is longer than the dependency between *how* and the verb. Thus, the reader releases *how* at the point of the VP, and hence the maintenance of *how* does not affect the reading times at the matrix verb. Our results show increased reading

[7] We also took one region prior to the critical region (matrix verb spillover region) as a predictor to check the covariance between the region before the prior region and the crititical region; the covariance did not affect the results (who_{obj} /*how* vs. *why*: Estimate: 0.02, SE= 0.01, t=2.20).

times at the matrix verb for *who*$_{obj}$, suggesting that integrating *who*$_{obj}$ with the verb is costly at the point of the matrix verb.

2.9. Maintenance of Different Kinds of WhPs: A Self-Paced Reading Experiment 2b

The current study examines the construction where *who* was in the subject position. When *who* is in the subject position (*who*$_{subj}$), we expect different complexity effects with *who* in the object position (*who*$_{obj}$). Specifically, at the middle of the relative clause, *who*$_{obj}$ is predicted to be read significantly slower than *who*$_{subj}$ and *why*. This is because *who*$_{subj}$ and *why* should be released from memory as soon as the parser encounters the embedded subject. Therefore, we expect the reading times for *who*$_{obj}$ to be significantly slower than *who*$_{subj}$ and *why* at the embedded NP position.

2.9.1. Participants, Materials and Design

For this experiment, 78 Northwestern University students who were native English speakers with no history of language disorders gave informed consent and participated. In exchange for their participation, the participants were granted 1 credit necessary for introductory linguistic classes at Northwestern.

Critical items consisted of 24 sentence in the form of a 1x4 within-subjects factorial design, in which different kinds of WhPs (*who*$_{obj}$, *who*$_{subj}$, *that*, *why*) were manipulated as independent factors. A sample set of stimuli is summarized in Table 15. To ensure that participants did not encounter the same types of target items consecutively, items were distributed in a pseudo-randomized manner. In addition to the current experimental items, there were 80 filler sentences that involved irrelevant manipulations to the current ones. The experiment took around 30 minutes to complete.

Table 15. Sample stimuli for experiment 2b.

Factors	
WhPs	Examples
who$_{obj}$	The father considered who the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to.
who$_{subj}$	Who considered that the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma?
that	The father considered that the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma.
why	The father considered why the naive babysitter that the spunky children loved unexpectedly and benevolently handed the toys to the grandma.

2.9.2. Procedure

The same procedure as in Experiment 2a was used.

2.9.3. Analysis

A similar analysis was employed as with Experiment 2a. For each region, the reading times slower than 1200ms were excluded from the analysis. Participants whose accuracy was below 72% were excluded from the analysis. The critical regions included the embedded NP critical region, and one word following the critical region (embedded NP spillover region). The second critical region included the matrix verb critical region, and the word following the matrix verb critical region (matrix verb spillover region).

2.9.4. Results & Discussion

Region-by-region reading times are presented in Figure 8, the graph at the critical region (embedded NP) in Figure 9, the graph at the spillover region in Figure 10, the graph at the matrix verb in Figure 11, the matrix verb spillover region in Figure 12, and mixed effect model outputs are presented in Table 16. Mean accuracy for critical trial comprehension questions was 79.0%.

Table 16. Summary of mixed effect model outputs for experiment 2b.

Region		Estimate (SE)	t-value
Embedded NP Critical Region	Intercept	5.88 (0.03)	214.36
	who_{obj}*/how/why vs. that*	-0.01 (0.01)	-0.60
	who_{obj} *vs.* who_{subj}*/why*	0.02 (0.01)	2.35 *
	who_{subj} *vs. why*	0.01 (0.01)	1.26
Embedded NP Spillover Region	Intercept	5.97 (0.03)	221.18
	who_{obj}*/how/why vs. that*	0.02 (0.01)	1.57
	who_{obj} *vs.* who_{subj}*/why*	0.02 (0.01)	1.86
	who_{subj} *vs. why*	0.01 (0.01)	0.80
Matrix Verb Critical Region	Intercept	5.88 (0.02)	241.70
	who_{obj}*/how/why vs. that*	0.01 (0.01)	1.40
	who_{obj} *vs.* who_{subj}*/why*	0.03 (0.01)	2.76 *
	who_{subj} *vs. why*	0.01 (0.01)	0.92
Matrix Verb Spillover Region	Intercept	5.85 (0.02)	275.04
	who_{obj}*/how/why vs. that*	-0.01 (0.01)	-0.77
	who_{obj} *vs.* who_{subj}*/why*	0.03 (0.01)	3.84 **
	who_{subj} *vs. why*	-0.01 (0.01)	-0.32

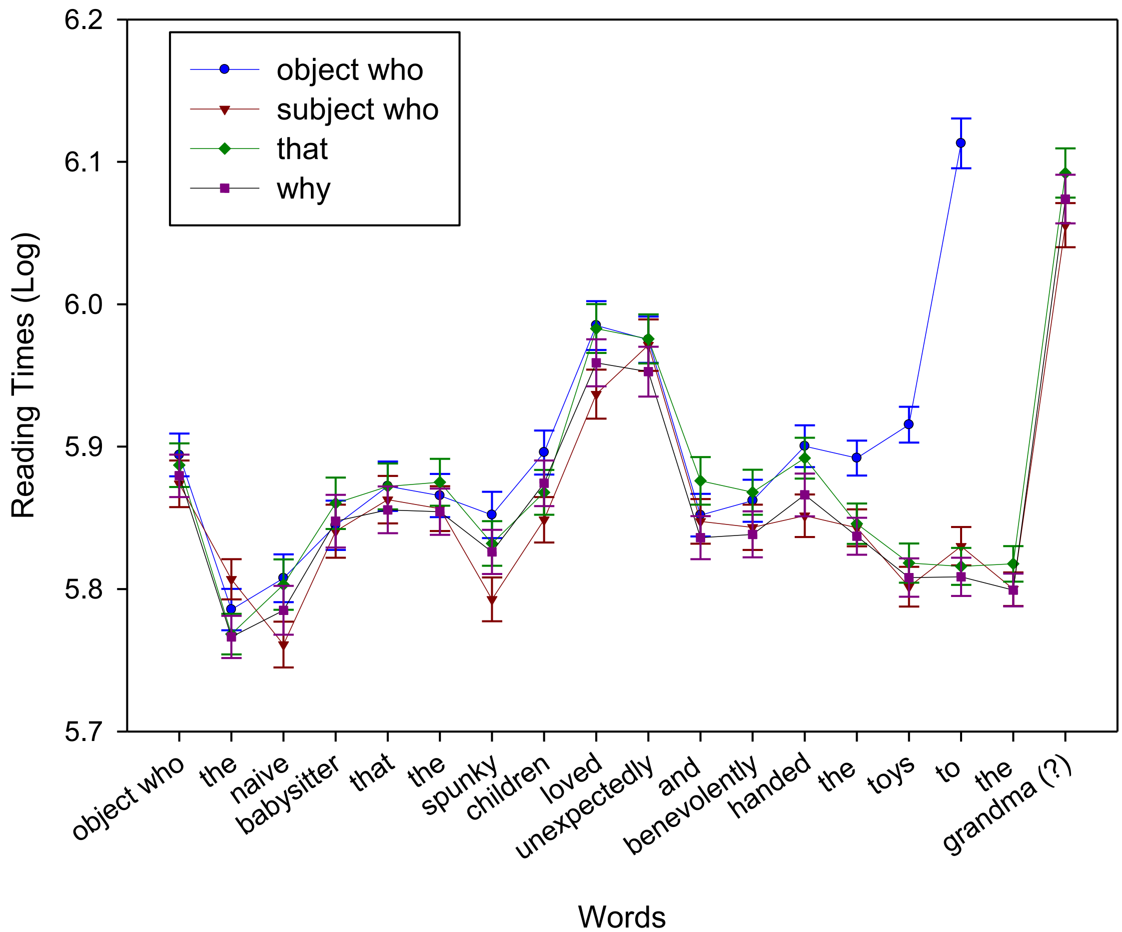


Figure 8: Region-by-region reading time means from experiment 2b. Note that the number of words at the beginning of the sentence differed according to the conditions. For example, it was either 'The father considered *who*/*how*/why…' for *who*$_{obj}$/*how*/*why* conditions, but 'Who considered that' for the *who*$_{subj}$ condition.

The results revealed that at the embedded NP, *children, who*$_{obj}$ was read significantly slower than *who*$_{subj}$ and *why*. At the embedded NP spillover region, a marginal significance was observed between *who*$_{obj}$ vs. *who*$_{subj}$/*why*.

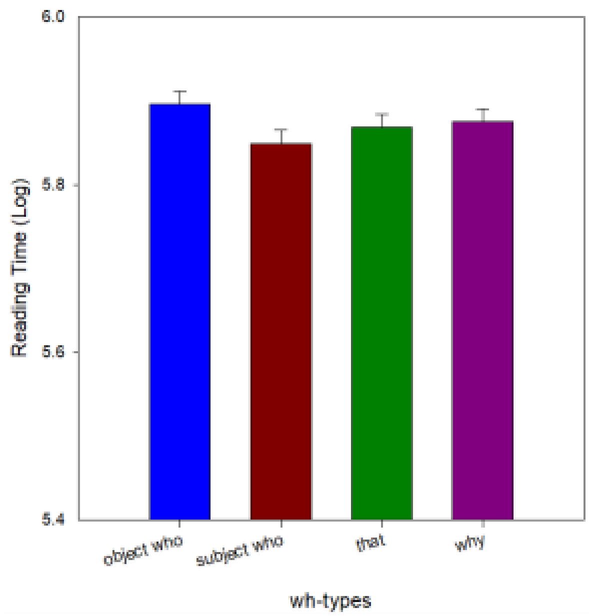


Figure 9. Reading times at the embedded NP region in experiment 2b.

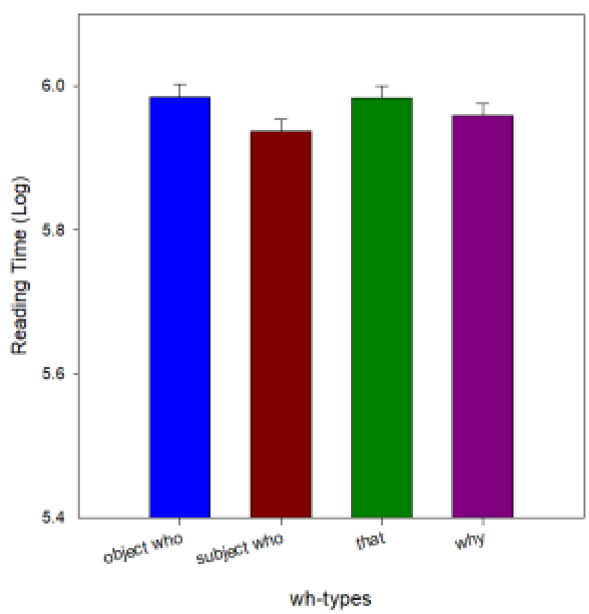


Figure 10: Reading times at the embedded NP spillover region in experiment 2b.

At the matrix verb spillover region, however, *who*$_{obj}$ was read signficantly slower than *why* and *who*$_{subj}$. At the matrix verb spillover region, *who*$_{obj}$ was read significantly slower than *who*$_{subj}$/*why*. This suggests that resolving the dependency is costly with the highest integration costs for *who*$_{obj}$ compared to *who*$_{subj}$ and *why*. In fact, at the preverbal adverb region 1 (*benevolently*), there was no sigificant effect, with no differences between *who*$_{obj}$ vs. *who*$_{subj}$/*why*.

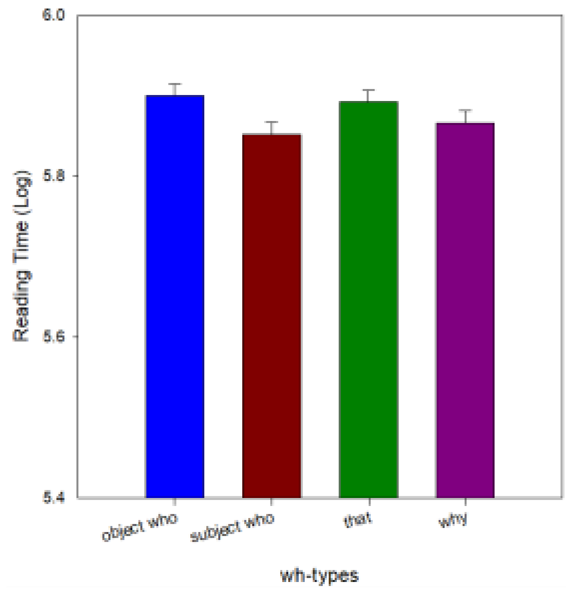


Figure 11: Reading times at the matrix verb in experiment 2b.

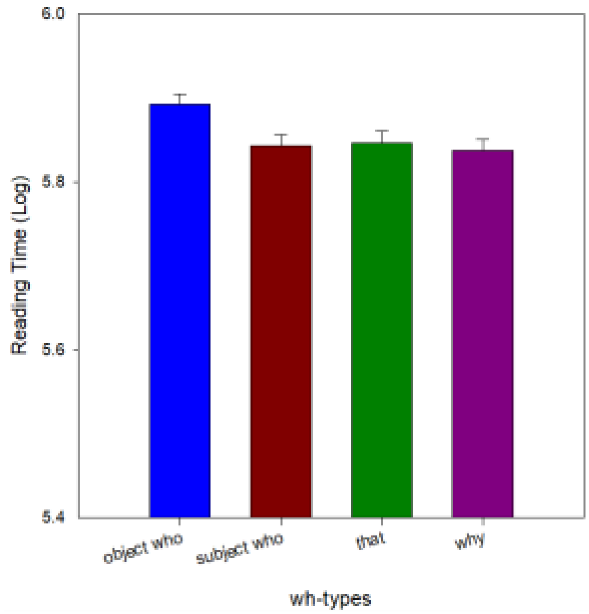


Figure 12: Reading times at the matrix verb spillover region in experiment 2b.

2.10. Interim Discussion

These studies are concerned with the storage component of the online WhFGD formation. Through online reading experiments and acceptability rating experiments, we showed that readers maintain the filler in memory where the maintenance cost is higher in who_{obj} and *how*, but not in who_{subj} and *why* at the deeply embedded noun. The acceptability rating experiment revealed that *why* is insensitive to negative islands whereas *how* is sensitive to negative islands. This provides further evidence that who_{subj} and *why* form a dependency with TP. However, some recent studies have suggested that *why* has two meanings, and the two different meanings (*why* asking for the reason, the reason *why* or why_{R}, and *why* asking for purpose, the purpose *why*, or why_{P}) correspond to two different structures. Chapman & Kučerová (2016) suggested that why_{R}, serving as a modifier for the proposition, is generated high and forms a dependency with TP, but *why_P*, serving as a modifier for VP, forms a dependency with VP (see Stepanov & Stateva, 2015, for related points). We tested whether two different *why*s in different syntactic structures give rise to different acceptability ratings in Q/A pairs: there was a significant difference between *why_R* and *why_P* with respect to the sensitivity to negative islands.

This created a good testing ground to examine the storage costs for different kinds of WhPs in a complex domain. We hypothesized that if different wh-elements are maintained until their licensors are encountered, and once licensed are released from memory, different wh-elements should lead to different processing costs associated with wh-phrases with different dependency lengths. In other words, WhPs should show processing differences in terms of the storage costs at the embedded NP, as different wh-phrases engage in different dependencies and have different dependency lengths.

We conducted two online experiments to understand how different WhPs are maintained in memory and whether they are released from memory once they are licensed by their licensors. Our first self-paced reading experiment revealed that who_{obj} and *how* behaved differently from *why* in terms of memory costs. Who_{obj} and *how* were read significantly slower than *why* at the embedded NP position. This suggests that once *why* is licensed by TP, it is released from memory, no longer impacting memory storage costs. However, who_{obj} and *how* need to be kept in memory until they reach the matrix verb, and hence increase processing complexity. At the matrix verb spillover region, there was a significant difference between who_{obj} vs. *how*/*why* suggesting increased integration costs for who_{obj} at the matrix verb.

The second experiment showed that who_{obj} behaved differently from who_{subj} and *why*. Given that the dependency length is the key predictor of whether the filler is released from memory or not, we showed that *who*$_{subj}$ and *why* are released early from memory once the dependency is formed. At the matrix verb, there was also a significant difference between *who*$_{obj}$ vs. *who*$_{subj}$/*why*. As *why* and *who*$_{subj}$ are released when TP is recognized, they do not increase complexity at the middle of the center-embedded RC. On the other hand, *who*$_{obj}$ and *how* must be held in memory until they reach V or VP. This suggests that an additional open dependency is made with the increased complexity at the middle of the center-embedded RC.

We argue that the reading time differences at the most deeply embedded NP suggest the differences of the storage costs for different WhPs. Under theories of processing complexity where processing complexity is dependent on the number of open dependencies, we expect the most deeply embedded NP position to have the highest storage costs. This is consistent with Gibson (1998)'s Syntactic Prediction Locality Theory (SPLT) where the WhFGD formation and the center-embedded RC interact in a way that the storage costs will be the highest at the the most deeply embedded NP position.

In our example, at the point of the embedded NP, *children*, there are different numbers of open dependencies in each WhFGD construction. If we assume that the number of open dependencies is correlated with processing complexity, then we could argue that the storage costs for *who*$_{obj}$ and *how* are higher at the embedded NP than *who*$_{subj}$ and *why*. For *who*$_{subj}$ and *why*, there are only two open dependencies waiting to be resolved, namely *the babysitter* and the verb as well as *the children* and the verb. On the other hand, for *who*$_{obj}$ and *how*, there are three open dependencies, namely *the babysitter* and the verb, *the children* and the verb, and *who*$_{obj}$ and the verb and *how* and VP.

In our online experiments, the reading times for *why* were faster than other wh-elements at the embedded NP. We assumed that *why* is located at Spec_CP, forming a dependency with TP. Therefore, there are only two open dependencies that wait to be resolved at the embedded NP. However, as Chapman & Kučerová notes, there can be two different *why*s, and one of which, *why_P*, serves as a modifier for VP, forms a dependency with VP. If the parser picks *why_P*, we do not expect any difference in reading times at the embedded NP. Note that our acceptability rating experiment tested the sensitivity to negative islands on *why_R* and *why_P*. The results of an acceptability rating experiment in 1d revealed a sensitivity to negative islands for *why_P* but not to *why_R*, suggesting that *why_P* moves from VP area to CP, i.e., forming a dependency with VP. Then why is *why* behaving differently from other wh-phrases like who_{obj}/*how* in online experiments if there are two different *why*s? We argue that *why* is lexically ambiguous between *why_R* and *why_P*, but the parser picks the shortest dependency driven by economic reasons (Gibson, 1998, 2000). This indicates that a dependency between a wh-element and TP is shorter than a dependency between a wh-element and VP. When the parser encounters TP, it picks the *why_R*, and releases this wh-phrase from memory. If we assume that the parser chooses the shortest dependency to reduce memory costs such as TP instead of VP in this case, it naturally follows why other wh-phrases are read slower at the deeply embedded NP than in *why* conditions.

These results are predicted if the WhPs are maintained in memory until they are linked to their licensors, but once linked to their licensors, they are released from memory. As Frazier (1985) and Gibson & Thomas (1999) suggest, if syntactic nodes are integrated and semantically combined, these nodes can be forgotten in a complex environment. Note that the notion of forgetting is different from what Frazier (1985) and Gibson &Thomas (1999) argue, as forgetting is more similar to 'forgetting from maintenance' (Wagers & Phillips, 2014). According to Frazier (1985) and Gibson & Thomas (1999), forgetting of the structure occurs when the memory load is high, in order to reduce the impact on memory resources. On the other hand, what we suggest is that the parser releases the structure when the dependency is completed, not only when the memory load is high. These results are compatible with the findings of Wagers & Phillips (2014) that certain information is maintained well (e.g., syntactic category information), whereas other information (e.g., semantic incompatibility) is less likely to be retrieved.

Finally, we would like to discuss the parsing mechanisms and possible parsing steps necessary to capture the results of this experiment. We assumed that different WhPs have different licensors and grammatical properties, where *why* is directly inserted in CP and licensed by TP, whereas *who* and *how* are licensed by the verb and the VP, respectively. However, from the perspective of online structure building, CP should already assume the existence of TP and VP due to the selectional relation and X' theory we are assuming. This suggests when CP is encountered, it already entails TP and builds the structure of T, and the presence of T entails VP. If this is the case, *why* and *how* which are licensed by the TP and VP should show similar processing profiles as they are both released from memory upon encountering TP. This is because when the parser encounters CP, the parser should be able to release *why* and *how* from memory at the same point. However, our data shows that *why* and *how* are processed differently in a direction based on the structural/linear distance between WhPs and their licensors. The parser should allow for the steps where there are parital trees not connected to bigger trees. In any case, if we assume that the parsing takes place from left to right, there should be a stack-like maintenance component (Frazier, 1987; Wanner & Maratos, 1978) and that there should be a state where *why* and TP exist, but VP does not exist. There are two possible ways to capture these differences in processing complexity between *why* and *how* in terms of parsing. The first way is to assume that the parser does not make use of grammatical information (e.g., selectional relation/subcategory feature). Even if the X' theory and selectional information entails the presence of the structure (i.e., TP entails VP), the parser would not use such information in the structure building process. An alternative way is to opt for the arc standard left corner parsing where it allows the partial structures to float around without integrating into the existing or bigger structure. Therefore, although CP and TP exist, TP will not dominate VP. This can be potentially implemented by assuming X' context free grammar as well as some kind of arc standard left corner parser (Resnik, 1992) to explain our experimental data.

2.11. Active vs. Reactivated Fillers

Many of the previous studies of wh-dependency processing have either adopted the maintenance view (Fiebach et al., 2002; Gibson, 1998; Gibson & Warren, 2004; Grodner & Gibson, 2005; Wagers & Phillips, 2014; Wanner & Maratsos, 1978; Warren & Gibson, 2002) or the retrieval view (Lewis & Vasishth, 2005; McElree, 2001, 2006; McElree & Dosher, 1989; Nicol et al., 1994; Nicol & Swinney, 1989; Van Dyke, 2007; Van Dyke & McElree, 2006). The maintenance and retrieval views are mostly motivated on empirical grounds. Storage cost effects (Chen et al., 2005; Gibson & Warren, 2004; Nakatani & Gibson, 2008) and active dependency formation (Phillips, 2006; Stowe, 1986; Traxler and Pickering, 1996), as reviewed earlier, provide motivation for the maintenance view. On the other hand, it has been shown that many effects attributed to storage cost can instead be understood as retrieval effects (Nicenboim et al., 2015). There are also some findings that are not compatible with the expectation-based (and storage) theories, such as their difficulty in predicting particular distance effects where facilitation is stronger for modifiers of the head of the dependency (Nicenboim, Logačev, Gattei, & Vasishth, 2016; Vasishth & Lewis, 2006). For example, (20b) should lead to the facilitation in reading times as there are more materials associated with VP (Vasishth & Lewis, 2006:776).

(20) a. that paper which that boy-ERG saw very old was. 'That paper which that boys saw was very old.'

b. that paper which that boy-ERG table-GEN behind fallen saw very old was. 'That paper which that boy saw fallen behind a/the table was very old.' (translation of German to English)

These observations have motivated the retrieval view. We argue that there is a possible mechanism that incorporates both retrieval and maintenance components, which has not been extensively investigated (Fiebach et al., 2002; Wagers & Phillips, 2014). In such a mechanism, maintained information is easier to access and unmaintained information is less accessible for retrieval when a gap is recognized. If some information associated with the filler is maintained and is less susceptible to decay, we expect it to be accessed easily (Wagers & Phillips, 2014).[8] On the other hand, if some information is susceptible to decay, we expect

---

[8] As an anonymous reviewer in *Language, Cognition, and Neuroscience* pointed out, Lewis & Vasishth (2005) suggests that retrieval could occur in such a way that the parser can re-instate information into comprehenders' focus of attention, in order to process that information. In this sense, if information were already in comprehenders' focus of attention

its retrieval to be more difficult. Another goal of the present study is to uncover the mechanism working behind both the maintenance and retrieval components by testing what aspects of a filler are retrieved in different WhFGD constructions: we refer to these as the *reactivated* WhFGD in (21a) and the *active* WhFGD in (21b).[9]

(21) a. *Reactivated Wh-Filler-Gap Dependency*

Which mistake in the program/programs __ will be disastrous for the company and certainly __ is/are harmful for everyone involved?

b. *Active Wh-Filler-Gap Dependency*

Which mistake in the program/programs [$_{RC}$ that will be disastrous for the company] certainly __ is/are harmful for everyone involved?

In (21a), the wh-filler must be linked to two gaps in the coordinate structure. When a sentence like (21a) is processed, the wh-filler is first linked to the gap in the first conjunct. Before the coordination connective *and* is encountered, the first conjunct can be understood as an independent sentence (*Which mistake in the program/s will be disastrous for the*

---

due to maintenance, there is no need for it to be retrieved. However, following Wagers & Phillips (2014), we argue that comprehenders discharge some components associated with the features from focal attention and this information must be retrieved when the verb is processed.

[9] Gaps are indicated by an underscore '__' in a sentence.

company?), thus the WhFGD can be resolved and interpreted at the point of the first gap. However, when the connective *and* is recognized, the wh-filler needs to be reactivated so that another WhFGD can be formed. This is so because the WhFGDs in the coordination construction obey grammatical constraints known as the Coordinate Structure Constraint (CSC) and the Across-the-Board (ATB) movement restriction (Ross, 1967). Specifically, wh-phrases cannot be extracted from only one conjunct in a coordinate structure, as a single conjunct in the coordinate structure is an island for wh-extraction (Ross, 1967). However, Ross (1967) has shown that wh-extraction from a conjunct is possible when the wh-phrase is extracted from all conjuncts. Thus, as shown in an example (22a), if any conjunct in a coordinate structure contains a gap, then all conjuncts must contain a gap, i.e., the wh-phrase needs to be extracted in an across-the-board (ATB) fashion (Gazdar, Klein, Pullum, & Sag, 1985; Ross, 1967; Williams, 1978). If not, the example is unacceptable, as (22b) shows.

(22) a. Which mistake __will be disastrous for the company and certainly __ is harmful for everyone involved?

b. *Which mistake __ will be disastrous for the company and certainly this mistake is harmful for everyone involved?

This suggests that in order to construct a grammatical WhFGD in a coordinated structure, the parser needs to posit the gap in the second conjunct subsequently to the first conjunct, and link the wh-phrase to the gap again in the second conjunct (see Wagers & Phillips, 2009 and Wagers & Phillips, 2014 for related experimental investigations). Thus, it should be the case that when the parser encounters the coordinating connective *and* the wh-phrase must be *reactivated* (*Reactivated Filler*).

On the other hand, (21b) involves a simple WhFGD construction. Although the wh-phrase is modified by a relative clause, the wh-verb dependency is established only at the main verb (the second verb *is/are*).[10] In (21b), the NP, which is the head of the relative clause, is linked to the gap within the relative clause. Thus, a filler-gap dependency is formed. However, unlike in (21a), the first half of the sentence (the wh-NP and the relative clause: *[$_{NP}$ Which mistake in the program [$_{RC}$ that will be disastrous for the company]]*) cannot be understood as an independent sentence. Furthermore, even though the head of the relative clause is linked to the gap within the relative clause, no WhFGD has been established at the point of the first gap position: the Wh-filler needs to be linked to the gap in the matrix clause for proper interpretation. Assuming that the parser engages in active dependency formation in a case like (21b), we call the wh-filler in (21b) the *Active Filler*.

If, as we have discussed earlier, active fillers are maintained in memory, then it means that they are immediately accessible to the parser to use in online structure building. This means that an active filler should be easier to access, compared to a reactivated filler, at the point of processing the verb. This is because reactivated fillers are released from memory and need to be retrieved when the gap or the verb is recognized. Thus, detailed information from reactivated fillers should be harder to access at the point of processing the verb and completing the WhFGD. Consider the difference between (21a) and (21b) from

[10] Note, the relative head needs to be linked to the embedded verb, but this is not relevant to the wh-gap dependency formation in terms of wh-question formation.

the perspective of online sentence processing. In (21a), the wh-phrase is linked to the gap in the first conjunct, meaning that the wh-gap dependency has been formed and the wh-filler no longer needs to be maintained. This may mean that the wh-filler can be released from memory and no longer impacts memory resources. Subsequently, when the coordinating connective is encountered, the wh-phrase would need to be reactivated. On the other hand, in (21b), the wh-phrase must be linked to the gap in the matrix clause directly. Therefore, the wh-phrase must be maintained until the gap is encountered. If the element that is maintained is retrieved more easily, then we expect that the information associated with wh-filler in (21b) will be retrieved more easily than in (21a).

2.12. How do we Approach Maintenance and Retrieval?

How can one examine maintenance and retrieval differences between active and reactivated fillers? The current work appeals to the agreement attraction effect, where the local noun (e.g., a noun other than the head) erroneously licenses agreement (Pearlmutter, Garnsey & Bock, 1999; Wagers et al., 2009, among many others). We use this as a probe to examine what aspects of the filler are retrieved.

One of the important features of agreement attraction is that it is sensitive to grammatical properties of the subject NP that triggers the erroneous agreement relation (Lago, Shalom, Sigman, Lau, & Phillips, 2015; Parker & Phillips, 2017; Tanner, Nicol, & Brehm, 2014; Wagers et al., 2009). When the number on the head noun and the verb mismatch, i.e., when grammatical agreement is not established (e.g., *the mistake in the programs *are*), then a clear interference effect from the local noun (*programs*) is typically present. This facilitation in ungrammatical conditions is often called an *Illusion of*

Grammaticality (Dillon, Mishler, Sloggett, & Phillips, 2013; Lago et al., 2015; Nicol, Forster, & Veres, 1997; Parker & Phillips, 2017; Pearlmutter et al., 1999; Tanner, Grey, & van Hell, 2017; Tanner et al., 2014; Thornton & MacDonald, 2003; Wagers et al., 2009). When the number of the head noun matches the number of the verb, i.e., when number agreement is grammatical (e.g., *the mistake in the programs is*), there is typically no interference observed from the local noun within the modifier (*programs*), though inhibitory effects are observed in some studies (Acuña-Fariña, Meseguer, & Carreiras, 2014; Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008; Jäger, Engelmann & Vasishth, 2017; Nicenboim, Vasishth, Engelmann, & Suckow, 2018; Pearlmutter et al., 1999).

These data suggest that when subject-verb agreement is computed, the parser first computes the agreement relation between the head noun and the verb, and only when this fails, the local noun embedded within the modifier phrase is retrieved. In other words, the initial stage of subject-verb agreement processing is guided by the grammatical structure of the subject NP, i.e., the parser identifies the head noun and specifically refers to its number information, not the number from other nouns embedded within the subject NP (Phillips, Lau, & Wagers, 2011; Kim, Brehm, & Yoshida, 2019). We use this aspect of agreement processing to investigate the extent to which the information on the NP is accessed. If only the category information is maintained and the details about the content of NP are released from the maintenance, then we expect no illusion of grammaticality. On the other hand, if detailed information about the NP (such as information about the head and the modifier) is maintained, then we expect an illusion of grammaticality to be present. With this *selective fallibility* aspect of parsing (Phillips, Wagers, & Lau, 2011) in mind, let us consider the processing of active and reactivated fillers.

If the active filler is less susceptible to memory decay, and full details about the wh-filler are maintained, we expect parser to be able to access detailed information about the filler when the verb is processed. For example, in (21b) the wh-phrase contains category information (NP), and the representation of the noun head (*mistake*) and the modifier phrases ([$_{PP}$ *in the programs*]). If maintenance of a wh-phrase leads to easier retrieval, all of these pieces of information may be retrieved. If this is the case, then an illusion of grammaticality effect should appear in active filler constructions.

The reactivated filler in (21a), on the other hand, is linked to the gap in the first conjunct, forming a dependency, meaning that the parser no longer needs to maintain the wh-filler. Thus, the wh-filler could be released from maintenance. Given that already-processed elements are susceptible to memory decay (Lewis & Vasishth, 2005; McElree et al., 2003), it is plausible that less detailed information about the filler will be retrieved at the second gap position in the second conjunct. Wagers & Phillips (2014) argued that lexical/semantic information is lost at a long distance. We could ask what other information is lost, and specifically whether the filler's internal structure remains at a long distance. If the filler is maintained, then the internal structure will be more available for the parser and can lead to an illusion of grammaticality effect. If not, only the category information will be available. If only the category of the filler is retrieved, this would lead simply to a grammatical mismatch effect without the illusion of grammaticality, and interference from the local noun regardless of whether the grammatical subject-verb agreement is established.

Specifically, differences in retrieval and maintenance indexed by the illusion of grammaticality effect are predicted for items involving *Reactivated WhFGD formation* (the filler is linked to the verb once and the wh-filler is reactivated later) and *Active WhFGD formation.*

(23) a. Which mistake in the program/programs will be disastrous for the company and certainly is/are harmful for everyone involved? (=21a)

b. Which mistake in the program/programs that will be disastrous for the company certainly is/are harmful for everyone involved? (=21b)

Both involve a complex wh-NP, composed of a head noun modified by a prepositional phrase (PP) containing another noun. In both, the wh-phrase serves as the subject of the first and second clause. For the subject-verb agreement dependency to be resolved, the number feature of the verb (i.e., *is*/*are*) in the second clause and the silent gap should agree; differences in processing at the verb in the second clause inform what is maintained versus needs reactivation.

If the parser needs to reactivate the wh-filler again in the second clause, we do not expect detailed information of the wh-NP to be accessible (this includes the internal structure, including category information and a representation of both the head and the modifier). Thus, when encountering a matrix verb that mismatches the number feature of the head noun, we expect a cost in the ungrammatical conditions, without any agreement attraction.

On the other hand, if information associated with the filler is maintained and thus not susceptible to decay, we expect information about the internal structure to be accessed more easily. The parser may maintain sufficiently-detailed information associated with the filler, including the representation of both the head and the modifier, until the wh-dependency is completed. When the parser encounters a matrix verb (e.g., *are*) that does not license the number feature of the head noun phrase (e.g., *mistake*), the parser could activate another noun that would fix the number mismatch. Thus, the ungrammatical matrix verb could be erroneously licensed by the local noun *programs,* consistent with the typical agreement attraction effect (Wagers et al., 2009) observed with overt subject noun phrases. If the wh-NP is sufficiently detailed to enable readers to make use of the head vs. non-head information, an agreement attraction effect is expected, and is predicted to be selective to ungrammatical conditions. As such, assuming that stronger maintenance leads to easier retrieval, we expect more detailed information about the filler to be retrieved in (23b) compared to (23a), leading to more agreement attraction for (23b) than for (23a).

2.13. Motivation of the Studies (Experiment 3, 4, & 5)

To address the question of what content is maintained and accessed at the gap, we directly compare the differences in agreement attraction between constructions that involve a relative clause (active filler) and active dependencies based on reactivation (reactivated filler). We conducted three acceptability rating experiments accompanied by three self-paced reading experiments.

The first two sets of experiments (experiment 3 & 4) serve the purpose of understanding the processing of the WhFGD within coordinated structures, in order to approach the question of what is maintained and what motivates the maintenance. The purpose of the first experiment is to examine what information is retrieved at the gap in the coordinated structure, testing the hypothesis that Wagers & Phillips (2014) held: in the reactivated filler constructions (i.e., the WhFGD in a coordinated structure), only coarse-grained information of the filler is retrieved (e.g., category information). Agreement attraction serves as a diagnostic for to what extent details about the wh-filler are accessible: If only coarse-grained information such as category is accessible, we expect no agreement attraction. On the other hand, if detailed information of the wh-filler, including the filler's internal structure, is accessible, we expect an illusion of grammaticality.

The second experiment (experiment 4) examines what motivates the maintenance of a filler. We compared coordinated structures that involve a wh-filler with ones that do not involve a wh-filler. In coordinated structures involving a wh-filler, like (23a), the reader can recognize that the gap should be inserted in the second conjunct upon encountering the coordinating connective (Wagers & Phillips, 2009). However, when no wh-element is included and when the subject of the sentence is a simple definite NP (e.g., *The mistake in the program/programs __ will be disastrous for the company and certainly__* is *harmful for everyone involved*), the presence of the filler-gap dependency is not signaled. Thus, the reader can recognize the movement structure only when the gap in the subject position of the second conjunct is recognized. The second experiment shows that there is indeed such a difference between a wh-phrase and a definite NP. This suggests that in wh-constructions, the wh-filler is reactivated and made more accessible for the parser at the point that the verb is processed. In other words, the wh-filler in the coordinated construction is initially released, but is subsequently reactivated and maintained again in memory. In the definite NP construction on the other hand, detailed information about the filler is not maintained, and thus needs to be retrieved at the verb, making it harder to access and leading to less agreement attraction and no illusion of grammaticality. We argue that, if both the wh-phrase and the definite NP were retrieved at the second verb position in the same way, then no such difference should be observed for the illusion of grammaticality effect.

The aim of the third experiment (experiment 5) is to examine the role of the maintenance associated with wh-fillers. In a reactivated filler, the wh-gap dependency is completed in the first conjunct, thus, the wh-filler is released from maintenance in the first conjunct. The recognition of the gap in the second clause triggers the retrieval of the wh-element. Assuming that the element released from the maintenance is subject to decay, the reactivated filler is not immediately accessible for the parser when the second verb is processed. On the other hand, in the active filler construction, the wh-filler is maintained in memory and thus it is immediately accessible for the parser when the second verb is processed. As a result, the prediction is stronger agreement attraction for the active filler than the reactivated filler, as the active filler is better maintained and likely to result in easier accessibility of more information about the internal structure, carried over a long distance.

2.14. WhFGD in a Coordinated Stucture: Experiment 3a: An Acceptability Rating Experiment

2.14.1. Participants, Materials and Design

Participants were 38 native speakers of English from Northwestern University with no history of reading disorders. All participants provided informed consent and received credit (1 credit/45 minutes) in an introductory Linguistics class.

32 critical items were arranged in a 2×2 within-subjects factorial design, in which *Local noun* (singular vs. plural) and *Grammaticality* (grammatical vs. ungrammatical) were manipulated as independent factors. A sample set of stimuli is summarized in Table 17. Items were distributed in a pseudo-randomized manner to make sure that participants did

not receive the same type of experimental items sequentially. One experimental item was excluded from the analysis due to a typographical error. Experimental items were combined with 98 filler sentences with manipulations irrelevant to the experimental items. The experiment took around 30 minutes to complete.

Table 17: Sample stimuli for experiment 3.

Factors		
Local noun	Grammaticality	Examples
Plural	*Grammatical*	Derek recalls which mistake in the programs will be disastrous for the company and certainly is harmful for everyone involved.
Plural	*Ungrammatical*	Derek recalls which mistake in the programs will be disastrous for the company and certainly are harmful for everyone involved.
Singular	*Grammatical*	Derek recalls which mistake in the program will be disastrous for the company and certainly is harmful for everyone involved.
Singular	*Ungrammatical*	Derek recalls which mistake in the program will be disastrous for the company and certainly are harmful for everyone involved.

2.14.2. Procedure

Stimuli were displayed on a desktop PC using the Linger software package (Rohde, 2003). For each stimulus, participants observed only one sentence on the screen until they pressed the button to move on. After each sentence, they selected a numbered button from 1 to 7, where 1 being totally unacceptable and 7 totally acceptable. Four practice items were presented before the actual experimental items. Participants were instructed that there were no right or wrong answers.

2.14.3. Analysis

Data were analyzed using an ordinal mixed-effects model performed with the *ordinal* package in R version 3.2.3 (Baayen, 2008; Baayen et al., 2008; Bates et al., 2014; Jaeger, 2008). A cumulative logit model was used instead of the linear model as the linear model assumes a continuous and unbounded dependent variable. Each model included simple difference sum-coded fixed effects of *Local noun* (singular vs. plural; contrasts -0.5 and 0.5) and *Grammaticality* (grammatical vs. ungrammatical; contrasts -0.5 and 0.5) and their interactions. The maximal random effects structure justified by the data was contained in all models (Barr et al., 2013), including random intercepts for participants and items and random slopes for fixed effects where they converged; the random effects that accounted for the least variance were removed in the case of non-convergence. See model tables for random effect structures.

2.14.4. Results & Discussion

The quantiles of residuals were relatively small and symmetrical about zero (Min: -3.26, Median: 0.06, Max=2.71). Mean acceptability scores are shown in Table 18 and in Figure 13, and the ordinal mixed effect model outputs are shown in Table 19.

Table 18: Mean acceptability scores for experiment 3a.
(Standard errors are in parentheses)

Factors		Average raw rating (SE)
Local *n*oun	Grammaticality	
Plural	*Grammatical*	4.54 (0.13)
Plural	*Ungrammatical*	4.33 (0.11)
Singular	*Grammatical*	4.77 (0.10)
Singular	*Ungrammatical*	3.99 (0.14)

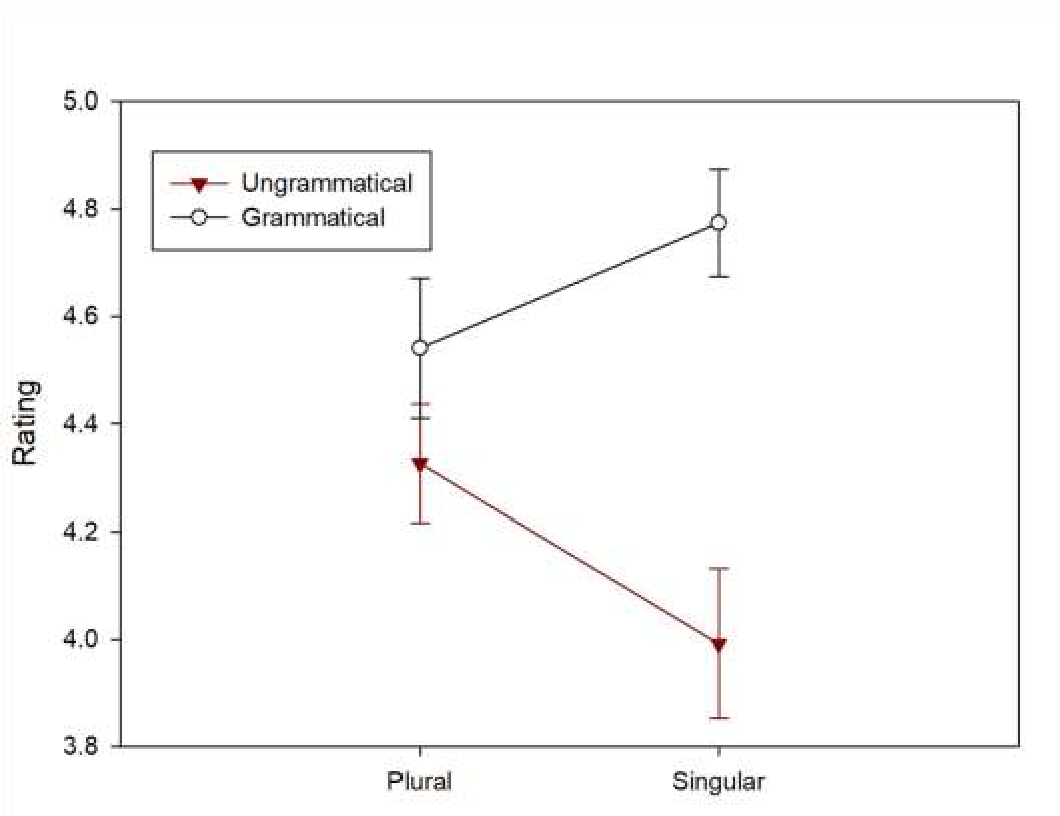


Figure 13. Mean acceptability scores for experiment 3a.

Table 19. Summary of fixed effects from the ordinal mixed effect model in experiment 3a.
Random intercepts were included for subjects and items, as were by-subject intercepts for *Local noun* and *Grammaticality*, and an interaction between *Local*

noun and *Grammaticality*, and by-item intercepts for *Local noun*, *Grammaticality* and an interaction between *Local noun and Grammaticality*.

	Estimate	SE	z	p
(Intercept)				
Local noun	*0.08*	*0.13*	*0.65*	*0.51*
Grammaticality	*-0.72*	*0.26*	*-2.74*	*<0.01 ***
Local noun x Grammaticality	*0.94*	*0.24*	*3.90*	*< 0.001****

We observed a significant main effect of *Grammaticality* where ungrammatical conditions were rated significantly less acceptable than grammatical conditions. This was qualified by an interaction between *Local noun* and *Grammaticality* where the difference between plural and singular local nouns was larger in ungrammatical conditions. This was further confirmed by a subset analysis, where the main effect of *Local noun* was larger in ungrammatical ($\beta = 0.52$, $SE= 0.19$, $z= 2.75$, $p < 0.01$) than in grammatical conditions ($\beta = -0.40$, $SE= 0.15$, $z= -2.59$, $p < 0.01$) and ungrammatical sentences were rated significantly less acceptable than grammatical sentences. This observed illusion of grammaticality provides evidence for the retrieval of grammatical properties, such as information about the internal structure such as the head and the modifier, in reactivated WhFGD constructions[11].

[11] As an anonymous reviewer from *Language, Cognition, and Neuroscience* pointed out, decay should have less impact on the offline experiments as readers can look back at the left context anytime, to remember the content of the antecedent. Our purpose of the offline experiments was to understand how the availability of the contexts can influence the retrieval of different kinds of information.

2.15. Experiment 3b: A Self-Paced Reading Experiment

2.15.1. Participants, Materials and Design

Participants were 58 native speakers of English from Northwestern University with no history of reading disorders. All participants provided informed consent and received credit (1 credit/ 45 minutes) in an introductory Linguistics class. Seven participants were excluded due to very low accuracy (<65%) in answering questions after each stimulus.

The same critical items were used as in Experiment 3a. Items were distributed in a pseudo-randomized manner to make sure that participants did not receive the same type of experimental items sequentially. Two experimental items were excluded from the analysis due to typographical errors. The experimental items were combined with 96 filler sentences of similar complexity. Fillers included items related to ambiguity resolution, passive sentences and locative constructions, all of which are irrelevant to processing either agreement attraction or coordinate structures.

2.15.2. Procedure

Stimuli were displayed on a desktop PC using the Linger software package (Rohde, 2003). A self-paced word-by-word moving window paradigm (Just et al., 1982) was employed. Participants saw a row of dashes, masking the words in the sentence. Participants pressed the space bar to proceed to the next sentence. After reading each sentence, they were asked to answer comprehension questions. To answer comprehension questions, participants were asked to press F (yes) or J (no) keys. An example comprehension question is *Was the word stadium mentioned in the story?*. They were provided with immediate feedback in terms of their accuracy. Six practice items were given to participants at the beginning of the experiment. The experiment took each participant about 30-45 minutes to complete.

2.15.3. Analysis

Data were analyzed using linear mixed effect regression, performed with the *lme4* package in R version 3.2.3 (Baayen, 2008; Baayen et al., 2008; Bates et al., 2014; Jaeger, 2008). Reading times were log-transformed to minimize non-normality (Box & Cox, 1964; Vasishth et al., 2013) and data that fell outside 2.5 standard deviations from the overall mean for the each region was excluded from the analysis. The critical regions are the verb, and the post-verb word comprises spillover region 1, which is then followed by the spill over region 2. The by-region exclusion percentages due to outlier removal were 1.73 % (verb region), 2.59 % (spillover region 1), and 1.5% (spillover region 2).

2.15.4. Results & Discussion

Figure 14 shows region-by-region reading times, Figure 15 shows the interaction plot at the critical region (spillover region 1), and Table 20 shows the mixed effect model outputs. Mean accuracy for critical trial comprehension questions was 78.0%.

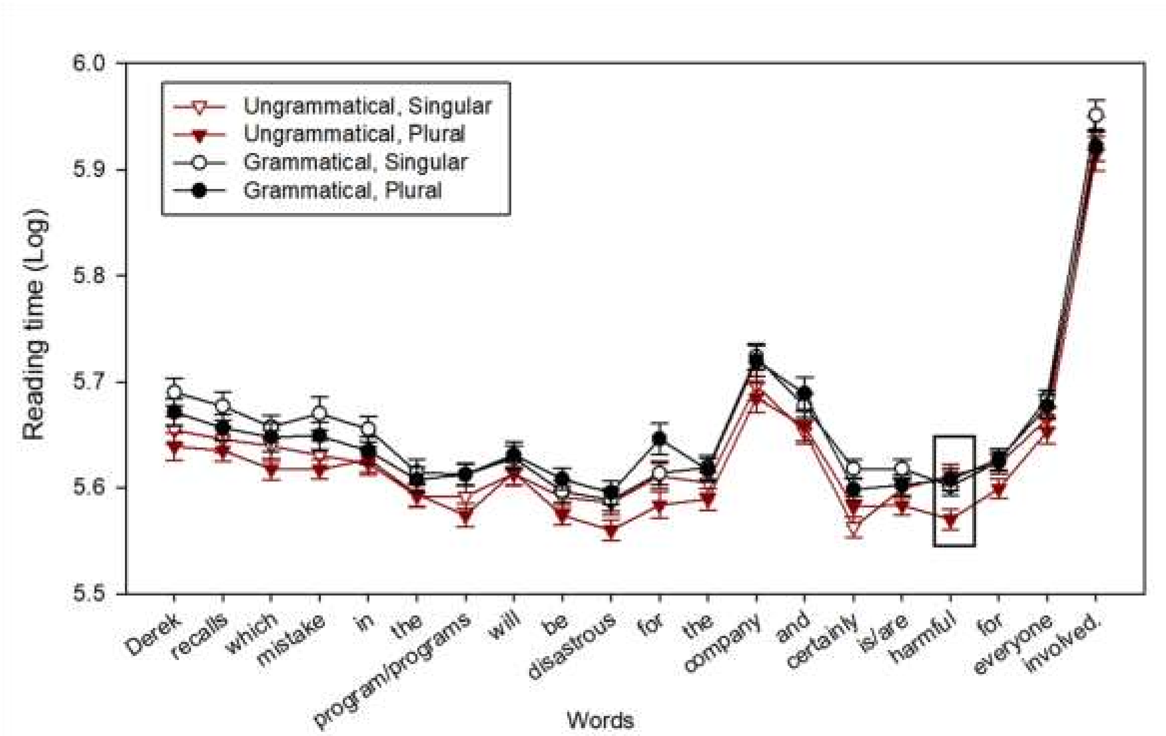


Figure 14. Region-by-region reading times for the experiment 3b. The box indicates the spillover region *harmful*.

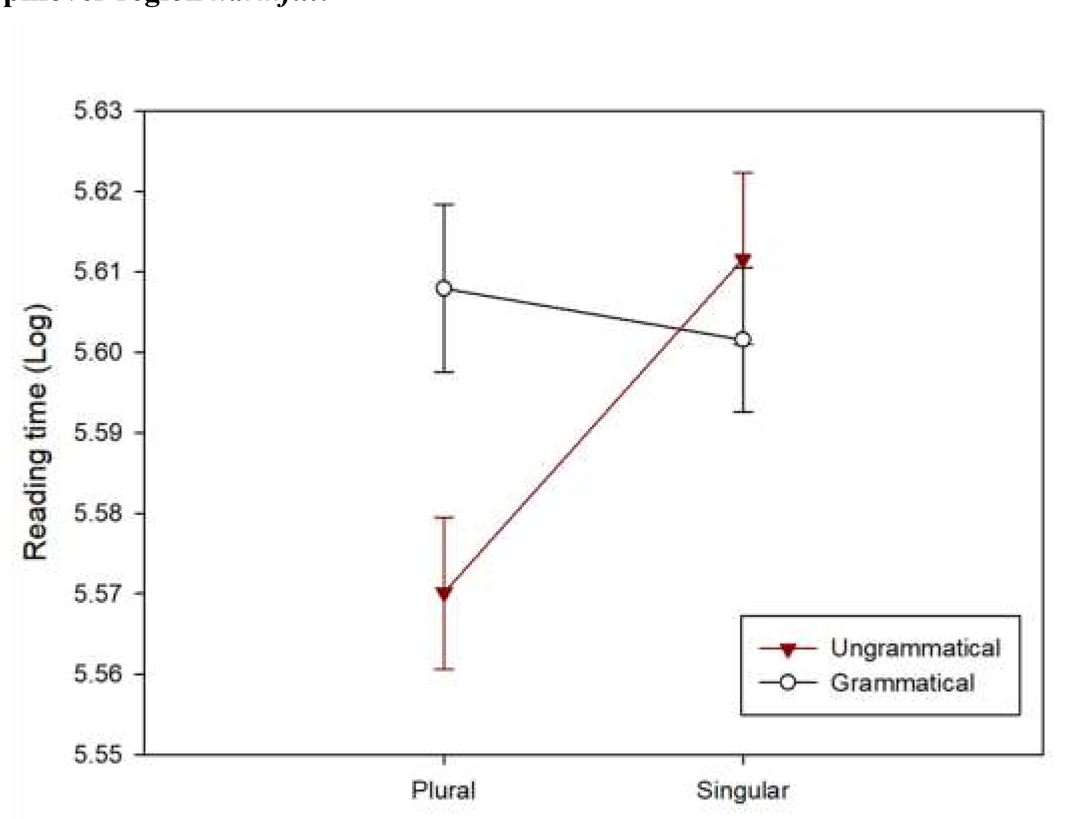


Figure 15. Interaction plot for spillover region 1 (*harmful*).

Table 20. 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 *Local noun* and *Grammaticality*, by-item random intercepts and slopes for *Local noun* and *Grammaticality*.				
(Intercept)	5.60	0.03	179.07	
Local noun	-0.02	0.01	-1.22	0.23
Grammaticality	-0.02	0.01	-1.40	0.17
*Local noun * Grammaticality*	0.00	0.02	0.08	0.93
Spill-over Region 1 (*harmful*): by-subject random intercepts and slopes for *Local noun*, *Grammaticality* and an interaction between *Local noun* and *Grammaticality*, by-item random intercepts and slopes for *Local noun*, *Grammaticality* and an interaction between *Local noun* and *Grammaticality*				
(Intercept)	5.60	0.03	182.99	
Local noun	-0.02	0.01	-1.55	0.13
Grammaticality	-0.01	0.01	-1.00	0.32
*Local noun * Grammaticality*	-0.05	0.02	-2.05	<0.05 *
Spill-over Region 2 (*for*): by-subject random intercepts and slopes for *Local noun*, *Grammaticality* and an interaction between *Local noun* and *Grammaticality*, by-item random intercepts and slopes for *Local noun* and *Grammaticality*, and an interaction between *Local noun* and *Grammaticality*.				
(Intercept)	5.62	0.03	185.24	
Local noun	-0.01	0.01	-1.07	0.29
Grammaticality	-0.02	0.01	-1.50	0.14
*Local noun * Grammaticality*	-0.02	0.02	-0.93	0.35

At the spillover region 1,[12] we observed a marginal main effect of *Local noun* where items paired with singular local nouns were read significantly slower than those with plural local

[12] As an anonymous reviewer in *Language, Cognition, and Neuroscience* has pointed out, it is possible that the absence of an agreement attraction at the verb might be due to the nature of the self-paced reading experiment. It has been well known that in self-paced reading experiments, the expected effect can be observed in one or two regions after the critical region (the spill-over effect; Vasishth & Lewis, 2006.). Therefore, it is possible that, even if the

nouns. We observed an interaction between *Local noun* and *Grammaticality* where constructions with singular local nouns were read slower than those with plural local nouns in ungrammatical conditions but no differences were detected in grammatical conditions. Subset analyses confirmed a main effect of *Local noun* ($\beta = -0.04, SE = 0.02, t = -2.77, p < 0.01$) in ungrammatical conditions, which was absent in grammatical conditions ($\beta = 0.01, SE = 0.02, t = 0.36$).[13] This again shows an illusion of grammaticality effect that provides evidence for the retrieval of grammatical properties in processing reactivated WhFGD constructions.

2.15.5. Discussion

We investigated what information associated with the filler is retrieved from memory in resolving reactivated WhFG dependencies. Ungrammatical sentences that included plural verbs resulted in high acceptability ratings as well as in decreased reading time, in

---

agreement attraction effect is caused at the verb region, it would not be observed right on the verb region but in spill-over regions.

[13] Following an anonymous reviewer's suggestion in *Language, Cognition, and Neuroscience*, we also examined the region immediately preceding the verb (i.e., the pre-critical region). The results showed a main effect of *Grammaticality* ($\beta = -0.03, SE = 0.01, t = -2.49, p < 0.05$) but no main effect of *Local noun* ($\beta = 0.00, SE = 0.01, t = 0.07, p > 0.05$) as well as no interaction between *Local noun* and *Grammaticality* ($\beta = 0.04, SE = 0.02, t = 1.49, p > 0.05$). This further suggests that the effects we observe are not due to spillover effects from the prior regions.

comparison to ungrammatical singular verbs, eliciting an illusion of grammaticality similar to that seen in overt sentences (Lago et al., 2015; Parker et al., 2017; Tanner et al., 2014). This suggests that grammatical information of the wh-filler is retrieved, including the representation regarding the head and the modifier (*which mistake in the programs*), allowing the verb to erroneously agree with the local noun as a last resort. In contrast, if detailed information associated with the wh-filler had not been recovered, all ungrammatical verbs would have been processed similarly, with no amelioration and reading time facilitation by a local plural noun.

It is possible that rather than a pure maintenance view, it is the presence of the coordinating connective *and* that triggers the reactivation and maintenance of the wh-filler and the active dependency formation. In other words, while retrieval happens at the gap, how much information is retrieved depends on how accessible the information is. The agreement attraction at spillover region 1 indicates that the grammatical and lexical content of the wh-NP are readily reactivated once the verb is processed. However, the lack of attraction at the verb region suggests that differences between conditions appear after processing the verb, and after processing the gap.

Our results are less compatible with the view that only the category information of the filler is accessible at the verb position. If only the category information were accessible, we would not expect agreement attraction to be present. The results are compatible with the view that the whole NP including category information (e.g., NP) and grammatical information (information about the internal structure; the representation regarding the head and the modifier) are retrieved, leading to an agreement attraction effect at the verb region.

2.16. Experiment 4a/4b: Wh-filler vs. Definite NP

Experiment 3 showed that readers retrieve detailed category and grammatical information, including the internal structure of the noun head and its modifier phrase. This led to an illusion of grammaticality effect. In the current experiment, we compare coordinated structures that involve a wh-filler, (24a) with those that do not involve a wh-filler, (24b).

(24) a. Which mistake in the program/programs __ will be disastrous for the company and certainly__ is harmful for everyone involved?

b. The mistake in the program/programs __ will be disastrous for the company and certainly__ is harmful for everyone involved.

One major difference between the two types of coordinated construction is that the former involves a wh-element that can signal the presence of the filler-gap dependency prior to encountering the gap.[14] Therefore, in the wh construction, the presence of the filler-gap

---

[14] An anonymous reviewer in *Language, Cognition, and Neuroscience* suggested that the recognition of the gap is not due to grammatical constraints such as the CSC and the ATB restriction. It could be the case that the readers recognize the presence of the gap due to the combination of the coordinating connective, *and*, and an adverb. If the combination of the coordinating connective and an adverb (... *and certainly* ...) helps reactivate the filler, then our assumption must be weakened, i.e., the reactivation of the filler is not due to the grammatical constraints. However, as Wagers & Phillips (2009) showed, the gap in the

dependency is recognized immediately upon encountering the wh-phrase and thus the parser can compute any grammatical constraints that apply to the WhFGDs such as CSC and the ATB restriction. If Wagers and Phillips (2009, 2014) are correct, then this means that the wh-filler can be reactivated upon encountering the coordinating connective *and*. On the other hand, the definite NP subject (e.g., *the mistake in the program/s*) does not signal the presence of a filler-gap dependency, and thus the coordinating connective should not reactivate the definite NP subject. As the presence of the filler-gap dependency is recognized when the gap in the second conjunct is recognized, the recognition of the gap and the retrieval of the subject NP in the first conjunct may occur at the same time. The prediction is that the definite NP subject should not be reactivated by the coordinating connective. Thus, retrieving a definite NP subject at the gap position in the second conjunct could be more difficult than retrieving the wh-filler, leading to a reduced agreement attraction effect.

---

coordinated structure and parasitic gap within an adjunct clause, which is optional, show different reactivation profiles. Therefore, it is still plausible that ATB/CSC plays a role in the reactivation of the wh-filler. As we do not have any evidence to distinguish the two hypotheses, we would like to leave this point open at this point.

2.17. Experiment 4a: An Acceptability Rating Experiment

2.17.1. Participants, Materials and Design

Participants were 39 native speakers of English from Northwestern University with no history of language disorders. All participants provided informed consent and received credit (1 credit/ 45 minutes) in an introductory Linguistics class.

32 critical items were arranged in a 2×2×2 within-subjects factorial design, in which *Local noun* (singular vs. plural), *Grammaticality* (grammatical vs. ungrammatical) and *Filler type* (the definite NP vs. wh-filler) were manipulated as independent factors. A sample set of stimuli is summarized in Table 21. Items were distributed in a pseudo-randomized manner to make sure that participants did not receive the same type of experimental items sequentially. The experimental items were combined with 56 filler sentences, with manipulations irrelevant to the current experiment. The experiment took around 30 minutes to complete.

Table 21. Sample stimuli for experiment 4.

Factors			
Local noun	Grammaticality	Filler type	Examples
Plural	*Grammatical*	*The Definite NP*	The mistake in the programs will be disastrous for the company and certainly is harmful for everyone involved.
Plural	*Ungrammatical*	*The Definite NP*	The mistake in the programs will be disastrous for the company and certainly are harmful for everyone involved.
Singular	*Grammatical*	*The Definite NP*	The mistake in the program will be disastrous for the company and certainly is harmful for everyone involved.
Singular	*Ungrammatical*	*The Definite NP*	The mistake in the program will be disastrous for the company and certainly are harmful for everyone involved.

Plural	*Grammatical*	*Wh-Filler*	Which mistake in the programs will be disastrous for the company and certainly is harmful for everyone involved?
Plural	*Ungrammatical*	*Wh-Filler*	Which mistake in the programs will be disastrous for the company and certainly are harmful for everyone involved?
Singular	*Grammatical*	*Wh-Filler*	Which mistake in the program will be disastrous for the company and certainly is harmful for everyone involved?
Singular	*Ungrammatical*	*Wh-Filler*	Which mistake in the program will be disastrous for the company and certainly are harmful for everyone involved?

2.17.2. Procedure

The similar procedure was employed as with Experiment 3a.

2.17.3. Analysis

A similar analysis was employed as in Experiment 3a. Each model included simple difference sum-coded fixed effects of *Local noun* (singular vs. plural; contrasts -0.5 and 0.5), *Grammaticality* (grammatical vs. ungrammatical; contrasts -0.5 and 0.5), *Filler type* (the definite NP vs. wh-filler; contrasts 0.5 and -0.5) and their interactions. The maximal random effects structure justified by the data was contained in all models (Barr et al., 2013), including random intercepts for participants and items and random slopes for fixed effects where they converged; the random effects that accounted for the least variance were removed in the case of non-convergence. See model tables for random effect structures.

2.17.4. Results & Discussion

The quantiles of residuals were relatively small and symmetrical about zero (Min: -3.40, Median: -0.04, Max=3.68). Mean acceptability scores are shown in Figure 16 and Table 22, and ordinal mixed effect model outputs are shown in Table 23.

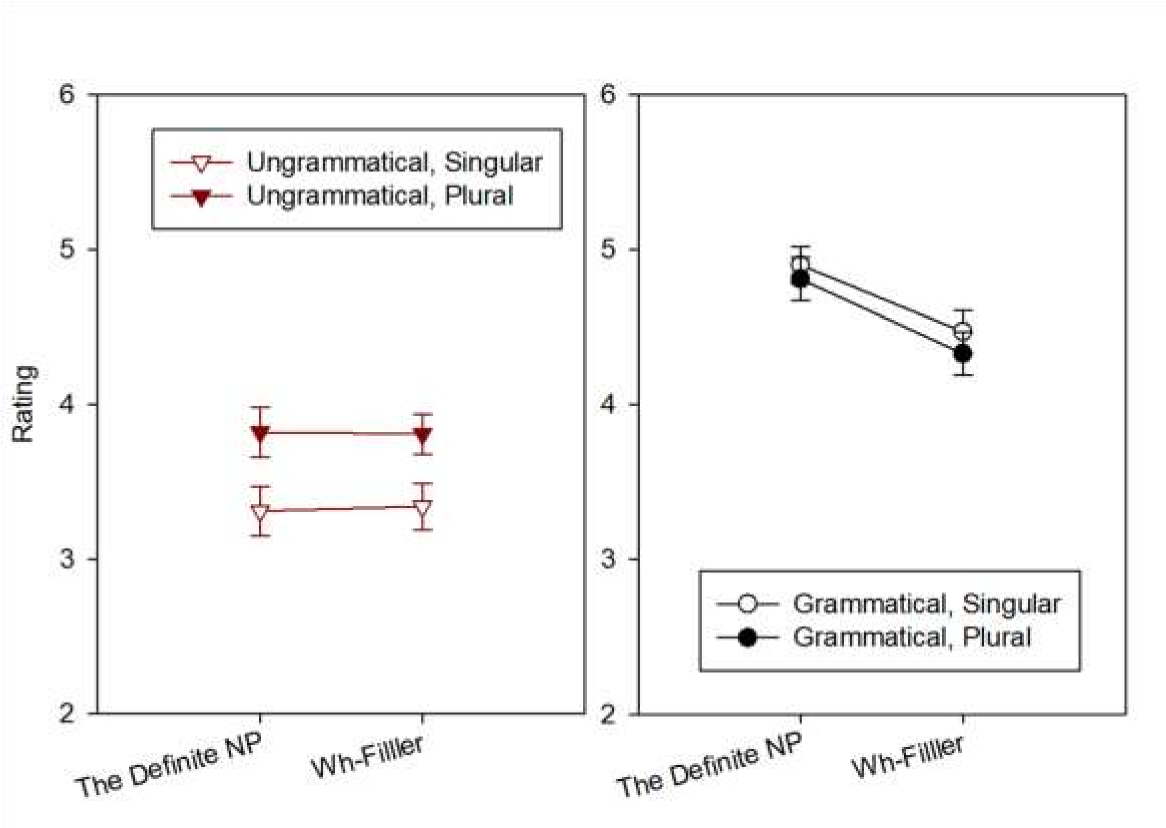


Figure 16. Mean acceptability scores for experiment 4a.

Table 22. Mean acceptability scores for experiment 4a.
(Standard errors are in parentheses)

Factors			
Local noun	Grammaticality	Filler type	Mean raw rating (SE)
Plural	*Grammatical*	*The Definite NP*	4.81 (0.14)
Plural	*Ungrammatical*	*The Definite NP*	3.82 (0.16)
Singular	*Grammatical*	*The Definite NP*	4.90 (0.12)
Singular	*Ungrammatical*	*The Definite NP*	3.31 (0.16)
Plural	*Grammatical*	*Wh-filler*	4.33 (0.14)
Plural	*Ungrammatical*	*Wh-filler*	3.81 (0.13)
Singular	*Grammatical*	*Wh-filler*	4.47 (0.14)
Singular	*Ungrammatical*	*Wh-filler*	3.34 (0.15)

Table 23. Summary of fixed effects from the ordinal mixed effect model in experiment 4a.

Random intercepts were included for subjects and items, as were by-subject intercepts for *Local noun, Grammaticality*, and *Filler type*, and by-item intercepts for *Local noun*, *Grammaticality*, and *Filler type.*

	Estimate	SE	z	p
(Intercept)				
Local noun	0.27	0.11	2.47	<0.05 *
Grammaticality	-1.74	0.32	-5.37	< 0.001***
Filler type	0.34	0.12	2.94	< 0.01**
Local noun x *Grammaticality*	0.92	0.21	4.39	< 0.001***
Grammaticality x *Filler type*	-0.63	0.21	-3.04	< 0.01**
Local noun x *Filler type*	0.12	0.21	0.56	0.57
Local noun x *Grammaticality* x *Filler type*	0.17	0.41	0.42	0.67

Local noun, *Grammaticality*, and *Filler type* were all significant as main effects. We found a main effect of *Local noun* where items paired with singular local nouns were rated lower than those with plural local nouns. We found a main effect of *Grammaticality* where ungrammatical items were rated as significantly less acceptable than those containing grammatical ones. Finally, a main effect of *Filler type* was observed, such that items with the wh-filler were rated as significantly less acceptable than those containing the definite NP.

We observed an interaction between *Local noun* and *Grammaticality* where constructions with singular local nouns were rated less acceptable than those containing plural local nouns, in the ungrammatical conditions only. This was further supported by subset analyses which confirmed a main effect of *Local noun* ($\beta = 0.71$, SE= 0.18, z= 3.87, $p < 0.001$) in ungrammatical conditions but not in grammatical conditions (*$\beta = -0.19$, SE= 0.15, z= -1.29, $p > 0.05$*). An interaction between *Filler type* and *Grammaticality* was also observed such that Definite NP filler types were judged to be significantly more acceptable than Wh-filler types in grammatical sentences only. This was confirmed with a subset analysis that revealed a main effect of *Filler type* (*$\beta = 0.70$, SE= 0.18, z= 3.87, $p<0.001$*) in grammatical conditions only. There were no interactions observed between *Local noun* and *Filler type*, or between *Local noun*, *Filler type*, and *Grammaticality*.

The pattern of increased acceptability for ungrammatical verbs following local plural nouns regardless of filler type indicates an illusion of grammaticality: ungrammatical definite NPs and Wh-Fillers are considered equally acceptable in offline ratings, despite the increase in acceptability for grammatical definite NPs over Wh-Fillers.

2.18. Experiment 4b: A Self-Paced Reading Experiment

2.18.1. Participants, Materials and Design

Participants were 81 native speakers of English from Northwestern University with no history of language disorders. All participants provided informed consent and received credit (1 credit/ 45 minutes) in an introductory Linguistics class. Six subjects were excluded due to their very low accuracy in answering comprehension questions about the sentences (<70%).

32 critical items were arranged in a 2×2×2 within-subjects factorial design, in which *Local noun* (singular vs. plural; contrasts 0.5 and -0.5), *Grammaticality* (grammatical vs. ungrammatical; contrasts -0.5 and 0.5) and *Filler type* (wh-filler vs. definite NP; contrasts 0.5 and -0.5) were manipulated as independent factors. Items were distributed in a pseudo-randomized manner to make sure that participants did not receive the same type of experimental items sequentially. The experimental items were combined with 56 filler sentences irrelevant to the current experiment.

2.18.2. Procedure

A similar procedure was employed as with Experiment 3b.

2.18.3. Analysis

The same factors and contrasts were used as in Experiment 4a. The rest of the analysis mirrored Experiment 3b, with the critical regions of the verb, the post-verb word (spillover region 1) and one word after the spill over region 1 (spillover region 2). The by-region exclusion percentages due to outlier removal were 1.43 % (verb region), 1.89 % (spillover region 1), and 1.74% (spillover region 2).

2.18.4. Results & Discussion

Region-by-region reading times for ungrammatical conditions are presented in Figure 17, the grammatical conditions presented in Figure 18, the interaction plot at the critical region in Figure 19, and mixed effect model outputs are presented in Table 24. Mean accuracy for critical trial comprehension questions was 83.0%.

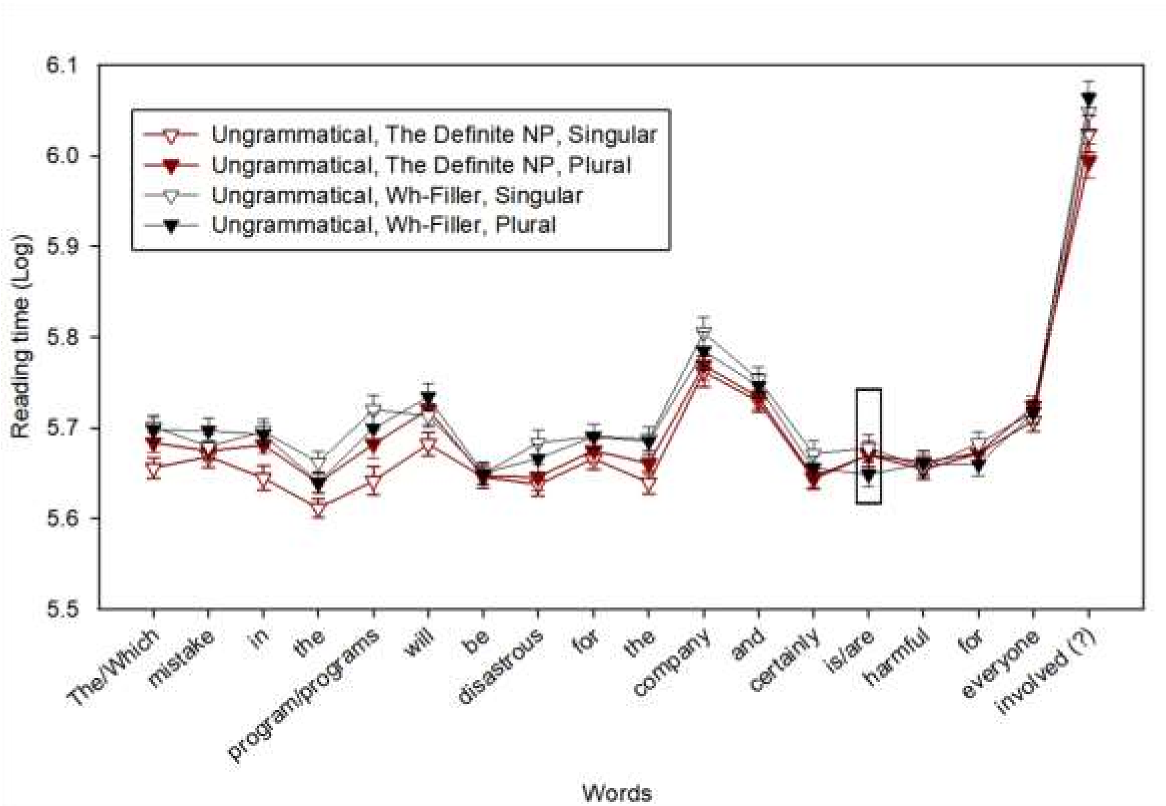


Figure 17. Region-by-region reading times for experiment 4b ungrammatical conditions. The box indicates the verb region *is/are*.

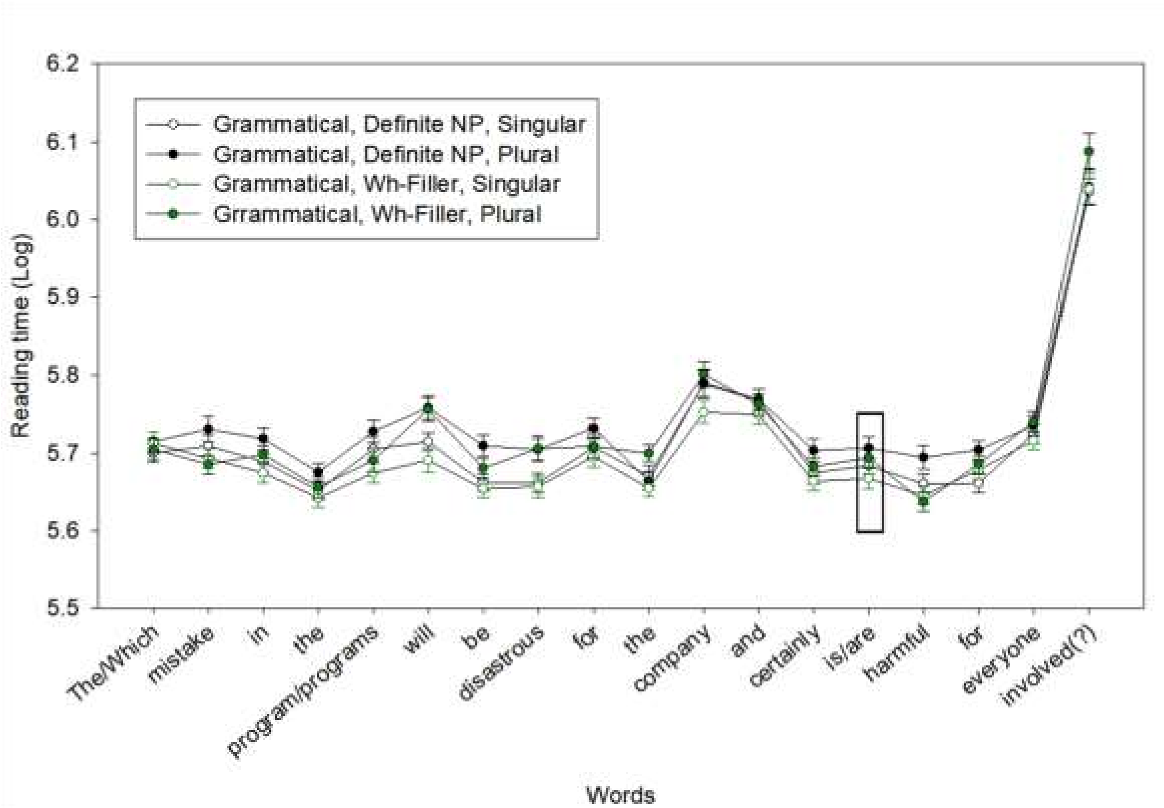


Figure 18. Region-by-region reading times for experiment 4b grammatical conditions. The box indicates the verb region *is/are*.

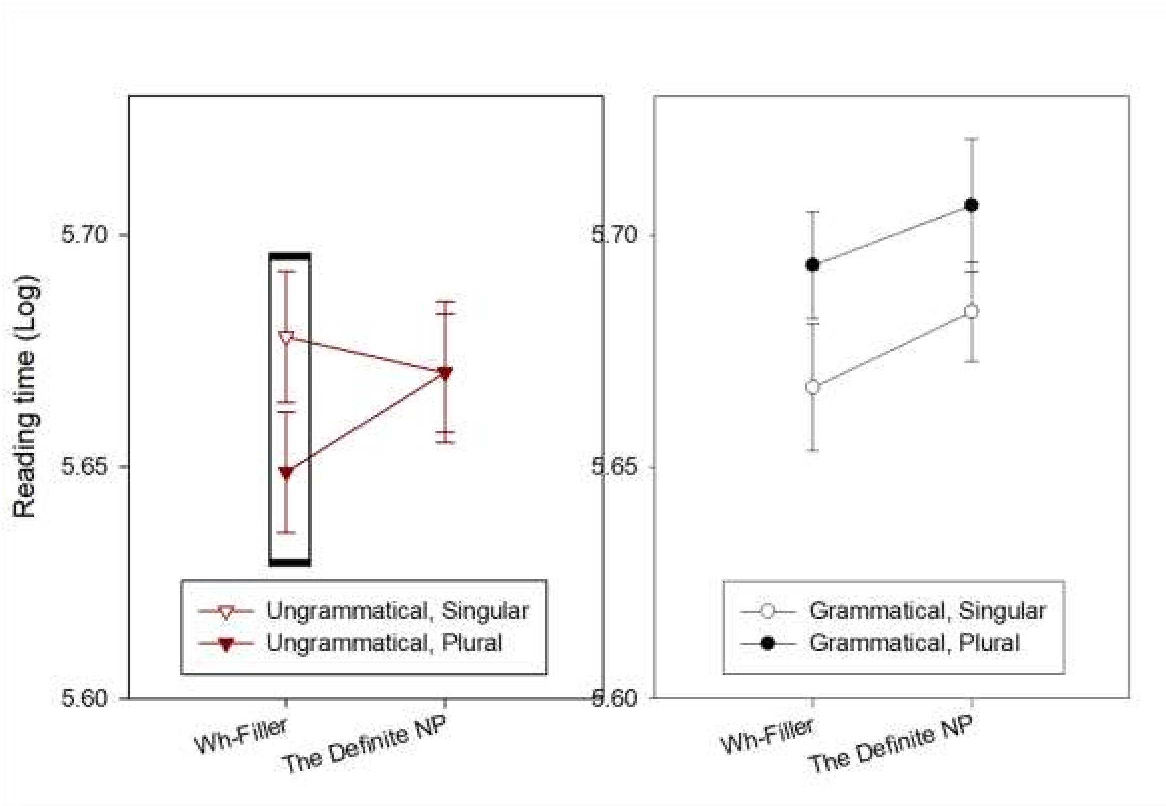


Figure 19. Interaction plot for critical verb region (*is/are*).

Table 24. Summary of results of linear mixed effects models by region in experiment 4b.

	Estimate	SE	t	p
Verb Region (*is/are*): by-subject random intercepts and slopes for *Local noun*, *Filler type* and *Grammaticality*, by-item random intercepts and slopes for *Local noun*, *Filler type* and *Grammaticality*				
(Intercept)	5.68	0.03	197.87	
Local noun	0.00	0.01	0.25	0.80
Grammaticality	-0.02	0.01	-2.21	<0.05 *
Filler type	0.01	0.01	1.16	0.25
Local noun x *Grammaticality*	-0.04	0.02	-2.18	<0.05 *
Grammaticality x *Filler type*	-0.01	0.02	-0.32	0.75
Local noun x *Filler type*	0.01	0.02	0.58	0.57
Local noun x *Grammaticality* x *Filler type*	0.04	0.03	1.13	0.26
Verb Spill-over Region 1 (*harmful*): by-subject random intercepts and slopes for *Local noun*, *Filler type* and *Grammaticality*, by-item random intercepts and slopes for *Local noun*, *Filler type* and *Grammaticality*				
(Intercept)	5.66	0.03	200.71	
Local noun	0.01	0.01	0.94	0.55
Grammaticality	-0.00	0.01	-0.10	0.99

Filler type	0.02	0.01	1.64	0.19
Local noun x *Grammaticality*	-0.01	0.02	-0.29	0.38
Grammaticality x *Filler type*	-0.03	0.02	-2.02	< 0.05 *
Local noun x *Filler type*	0.02	0.02	1.32	0.23
Local noun x *Grammaticality* x *Filler type*	-0.04	0.03	-1.06	0.44
Verb Spill-over Region 2 (*for*): by-subject random intercepts and slopes for *Local noun*, *Filler type* and *Grammaticality*, and by-item random intercepts and slopes for *Local noun*, *Filler type* and *Grammaticality*, and an interaction between *Grammaticality* and *Filler type*.				
(Intercept)	5.68	0.03	207.98	
Local noun	0.01	0.01	0.55	0.58
Grammaticality	-0.01	0.01	-1.18	0.24
Filler type	0.00	0.01	0.13	0.9
Local noun x *Grammaticality*	-0.04	0.02	-2.27	<0.05 *
Grammaticality x *Filler type*	0.00	0.02	0.07	0.94
Local noun x *Filler type*	0.03	0.02	1.75	0.08
Local noun x *Grammaticality* x *Filler type*	-0.00	0.03	-0.08	0.93

At the verb region, we found a main effect of *Grammaticality* where grammatical constructions were read slower than the ungrammatical constructions. This was driven by the critical interaction between *Local noun* and *Grammaticality*. A planned subset analysis showed that this interaction between *Local noun* and *Grammaticality* was significant only in the wh-filler NP ($\beta = -0.06, SE= 0.02, t= -2.33, p<0.05$) but not in the definite NP ($\beta = -0.02, SE= 0.03, t= -0.77, p>0.05$), indicating that the illusion of grammaticality was at least numerically driven by the reactivated wh-filler conditions, although the three-way interaction failed to reach significance.

At the spillover region 1, an interaction between the *Grammaticality* and the *Filler type* was observed such that the differences between the definite NP and the wh-filler were larger in grammatical conditions ($\beta = 0.03, SE= 0.01, t= 2.60, p<0.05$), indicating that the definite NP was read significantly slower than the wh-filler in grammatical conditions.

At the spillover region 2, we observed the critical interaction between *Local noun* and *Grammaticality* where the differences between plural local nouns and singular local nouns were larger in the grammatical conditions. A subset analysis confirmed that this was carried by a marginal main effect of *Local noun* in grammatical conditions ($\beta = 0.03, SE= 0.01, t= 1.83$). A marginal interaction between *Local noun* and *Filler type* was also observed. Further subset analysis revealed no main effect of *Local noun* in the wh-filler ($\beta = -0.01, SE= 0.02, t= -0.52$) but a marginal main effect of *Local noun* in the definite NP ($\beta = 0.02, SE= 0.01, t= 1.63$). This indicates that the singular local noun was read faster than the plural local noun in the definite NP.

2.18.5. Discussion

In this experiment, we tested whether coordination leads wh-NPs and definite NPs to be reactivated similarly at the gap (the verb) in the second conjunct. Although the three-way interaction did not reach significance, the results of the planned subset analysis are compatible with the idea that attraction was reduced for definite conditions relative to wh-fillers, suggesting that details about the grammatical information of the definite NP might not be retrieved at the verb. Assuming that this is correct, we argue that these differences in attraction are due to differences in how these two kinds of fillers are processed. While the wh-filler should be reactivated at the coordinating connective and put into maintenance again, this should not occur for the definite NP. This can be understood by considering the time-course of processing the verb in the second conjunct.

In the definite NP condition, when the reader encounters the coordinating connective *and*, the parser may expect a clausal conjunct which involves an overt subject and a verb, or another NP, due to the local attachment bias (Staub & Clifton, 2006). If so, when the parser encounters the verb, the parser needs to abandon this expected structure and build a structure with a subject gap. Because the gap is not expected upon encountering the coordinate structure, the parser could posit a gap only after the bottom-up evidence (the verb) is encountered leading to a reanalysis. In other words, in definite NP sentences, the definite NP itself does not signal the presence of a filler-gap dependency and the coordinating *and* does not provide a cue to actively complete the dependency: the parser does not maintain the definite NP subject. The lack of a significant illusion of grammaticality in the definite NP conditions is plausibly due to the fact that the information associated with the definite NP was not maintained and thus is subject to memory decay. At the same time, this may be due to the reanalysis difficulty that we have mentioned above. In other words, the reanalysis processes and the reactivation might happen at the same point (at the verb), and thus we may not be able to observe the effect of reactivation or the reanalysis effect could hide the reactivation effect.

In contrast, the presence of a significant interaction indicating the illusion of grammaticality in the wh-filler conditions suggests that the detailed information from the wh-filler was readily accessible at the second verb position. This observation leads to the following conclusions. First, it is possibly the case that grammatical constraints such as ATB movement restriction and CSC in the coordinate structures could lead the parser to the formation of the wh-dependencies in the second conjunct. If the parser is sensitive to the ATB restrictions, upon encountering the coordinating connective *and*, the parser would be sensitive to the constraints on WhFGD formation in the context of coordinate structures, such as CSC and the ATB restriction (Wagers & Phillips, 2009). These constraints lead to actively searching for the gap in the second conjunct (Wagers & Phillips, 2009, 2014), which could lead to the more robust illusion of grammaticality effect in the wh-filler condition.

Assuming that there is a genuine processing difference between the wh-construction and the definite NP construction, then this suggests that the combination of the wh-filler and the coordinate structure is crucial. This means either that the wh-filler should be affected by the presence of the coordinating connective or that the processing of the wh-construction does not involve the reanalysis process that would mask the illusion of grammaticality effect. If the lack of the illusion of grammaticality effect in the definite NP constructions is due to a lack of reactivation of the definite NP, then the presence of the illusion effect in the wh-filler construction should be due to the reactivation of the wh-filler by the coordinating connective. On the other hand, if the lack of the illusion effect in the definite NP construction is due to reanalysis (the parser initially expected an NP-conjunct after the coordinated connective and had to change the structure to the clausal conjunct with a gap), then, in the wh-filler construction, such reanalysis process should not have taken place. We contend that the reanalysis hypothesis predicts that the adverb or the verb in the second conjunct should be read slower in the definite NP conditions than in the Wh-filler conditions because the adverb or the verb disambiguate the structure and therefore trigger reanalysis. As has been long known, reanalysis incurs a processing cost (Schneider & Phillips, 2001; Sturt, Pickering, Scheepers, & Crocker, 2001). Therefore, if reanalysis takes place in the definite NP conditions, masking the agreement attraction effect, then we expect slower reading of the verb and/or the adverb in the second conjunct in the definite NP conditions than in the Wh-filler conditions. In our data, this effect was not observed, and there was no main effect of *filler type* in either region (Adverb: $\beta = -0.00, SE= 0.01, t= -0.36,$ $p>0.05$; Verb: $\beta = 0.01, SE= 0.01, t= 1.16,$ $p>0.05$). This suggests against the reanalysis hypothesis. Therefore, we conclude that it is more likely that the wh-phrase is reactivated at the connective position and put into maintenance again.

In contrast to the pattern observed in the online data, note that in the offline rating experiment (Experiment 4a), we observed clear evidence for agreement attraction in the definite NP as well as the wh-filler conditions. We argue that this discrepancy may arise from the availability of the contexts for the offline rating experiment: readers had more time to go back and read the first conjunct in the rating experiment, leading to an agreement attraction effects. For the online experiment (Experiment 4b), we argue that the parser recognizes these grammatical constraints in real-time, leading to an expectation of the upcoming gap position upon encountering the coordinating connective, and actively linking the wh-filler and the subsequent gap site (Wagers & Phillips, 2009). Thus, when the reader encounters the connective and the verb sequence, the parser could readily reactivate the wh-filler and the wh-filler is maintained. If some information about the filler is more accessible and less susceptible to decay, we expect information to be retrieved easily (Wagers & Phillips, 2014). Thus, reactivation at the coordinating *and* could suggest that the parser retrieves detailed information at the verb. This could lead to retrieval of fine-grained information at the gap, such that the plural local noun is read faster than the singular local noun in ungrammatical conditions.

Another possibility for the differences between the retrieval of the definite NP and the wh-filler is that they could behave differently in terms of encoding. Wh-words could be intrinsically more prominent than the definite NP because they have special morphology, function and semantics (Jäger et al., 2017). Although this is indeed a possibility, we have to note that it is difficult to distinguish the effects of prominence from maintenance.

2.19. Experiment 5a/5b: Active Filler vs. Reactivated Filler

The results of the previous experiments showed that reactivation of fillers could not be the sole cause of agreement attraction. In this experiment, we ask how active versus reactivated wh-fillers may differ in processing. We compare how the information retrieved at the matrix verb (*is/are*) could differ by changing the dependency configuration as in (25).[15]

(25) a. Which mistake in the program/programs will be disastrous for the company and certainly is harmful for everyone involved? (=21a)

---

[15] As an anonymous reviewer in *Language, Cognition, and Neuroscience* pointed out, there is an alternative explanation for Experiment 5 that would not rely on the reactivated vs. active distinction, but rather, on differences in cue-based retrieval. In (21b), the attachment site of the RC is actually ambiguous, such that *that will be disastrous* could modify either *mistake* or *program(s)*. If readers prefer to attach the RC low, to *program(s)*, then according to cue-based retrieval this noun phrase will be reactivated, rendering it more active in memory. This would yield stronger attraction rates at the main verb (*is*/*are*), since the local noun will have higher activation (and thus interfere more) in (21b) than in (21a). However, if the attachment of RC modulates the accessibility of the lower noun, we also predict a similarity-based interference effect. In other words, the local noun should be more accessible across-the-board and thus should give rise to an interference effect whether the agreement is grammatical or ungrammatical. This should not predict the illusion of grammaticality we observed, but rather an agreement attraction effect in both grammatical and ungrammatical conditions.

b. Which mistake in the program/programs that will be disastrous for the company certainly is harmful for everyone involved? (=21b)

As we noted earlier, (25a) involves the parser positing a gap in the second conjunct subsequent to the first conjunct, and linking the wh-phrase to the gap in the second conjunct. This indicates that when the parser encounters the coordinating connective *and*, the wh-phrase must be reactivated. If the release from maintenance is subsequently followed by retrieval of decayed information, then we expect that the wh-filler will not be immediately accessible for the parser. This would suggest weaker agreement attraction. Conversely, (25b) involves an active filler where the wh-filler needs to be maintained until the matrix verb in order to resolve the dependency. If the parser could avoid the release from maintenance, we expect that detailed information associated with the wh-filler will be accessible for the parser, leading to stronger agreement attraction in (25b).

2.20. Experiment 5a: An Acceptability Rating Experiment

2.20.1. Participants, Materials and Design

Participants were 43 native speakers of English from Northwestern University with no history of language disorders. All participants provided informed consent and received credit (1 credit/ 45 minutes) in an introductory Linguistics class.

32 critical items were arranged in a 2×2×2 within-subjects factorial design, in which *Local noun* (singular vs. plural) and *Grammaticality* (grammatical vs. ungrammatical) and *Dependency type* (Active Filler vs. Reactivated Filler) were manipulated as independent factors. A sample set of stimuli is summarized in Table 25. Items were distributed in a

pseudo-randomized manner to make sure that participants did not receive the same type of experimental items sequentially. The experimental items were combined with 64 filler sentences, irrelevant to the current experiment. The experiment took around 30 minutes to complete.

Table 25. Sample stimuli for the experiment 5.

Factors			
Local noun	Grammaticality	Dependency type	Examples
Plural	*Grammatical*	*Active Filler*	Which mistake in the programs that will be disastrous for the company certainly is harmful for everyone involved?
Plural	*Ungrammatical*	*Active Filler*	Which mistake in the programs that will be disastrous for the company certainly are harmful for everyone involved?
Singular	*Grammatical*	*Active Filler*	Which mistake in the program that will be disastrous for the company certainly is harmful for everyone involved?
Singular	*Ungrammatical*	*Active Filler*	Which mistake in the program that will be disastrous for the company certainly are harmful for everyone involved?
Plural	*Grammatical*	*Reactivated Filler*	Which mistake in the programs will be disastrous for the company and certainly is harmful for everyone involved?
Plural	*Ungrammatical*	*Reactivated Filler*	Which mistake in the programs will be disastrous for the company and certainly are harmful for everyone involved?
Singular	*Grammatical*	*Reactivated Filler*	Which mistake in the program will be disastrous for the company and certainly is harmful for everyone involved?
Singular	*Ungrammatical*	*Reactivated Filler*	Which mistake in the program will be disastrous for the company and certainly are harmful for everyone involved?

2.20.2. Procedure

A similar procedure was employed as in Experiment 3a.

2.20.3. Analysis

The same analysis was employed as Experiment 3a. Each model included simple difference sum-coded fixed effects of *Local noun* (singular vs. plural; contrasts: -0.5 and 0.5), *Grammaticality* (grammatical vs. ungrammatical; contrasts: -0.5 and 0.5), *Dependency type* (active filler vs. reactivated filler; contrasts: 0.5 and -0.5) and their interactions. The maximal random effects structure justified by the data was contained in all models (Barr et al., 2013), including random intercepts for participants and items and random slopes for fixed effects where they converged; the random effects that accounted for the least variance were removed in the case of non-convergence. See model tables for random effect structures.

2.20.4. Results & Discussion

The quantiles of residuals were relatively small and symmetrical about zero (Min: -3.06, Median: -0.01, Max=3.55). Mean acceptability scores are shown in Table 26 and Figure 20, and ordinal mixed effect model outputs are shown in Table 27.

Table 26. Mean acceptability ratings from experiment 5.

Factors			
Local noun	Grammaticality	Dependency type	Mean raw rating (SE)
Plural	*Grammatical*	*Active Filler*	3.71 (0.15)
Plural	*Ungrammatical*	*Active Filler*	3.24 (0.10)
Singular	*Grammatical*	*Active Filler*	3.71 (0.12)
Singular	*Ungrammatical*	*Active Filler*	2.81 (0.11)
Plural	*Grammatical*	*Reactivated Filler*	4.37 (0.14)
Plural	*Ungrammatical*	*Reactivated Filler*	3.73 (0.12)
Singular	*Grammatical*	*Reactivated Filler*	4.39 (0.14)
Singular	*Ungrammatical*	*Reactivated Filler*	3.19 (0.14)

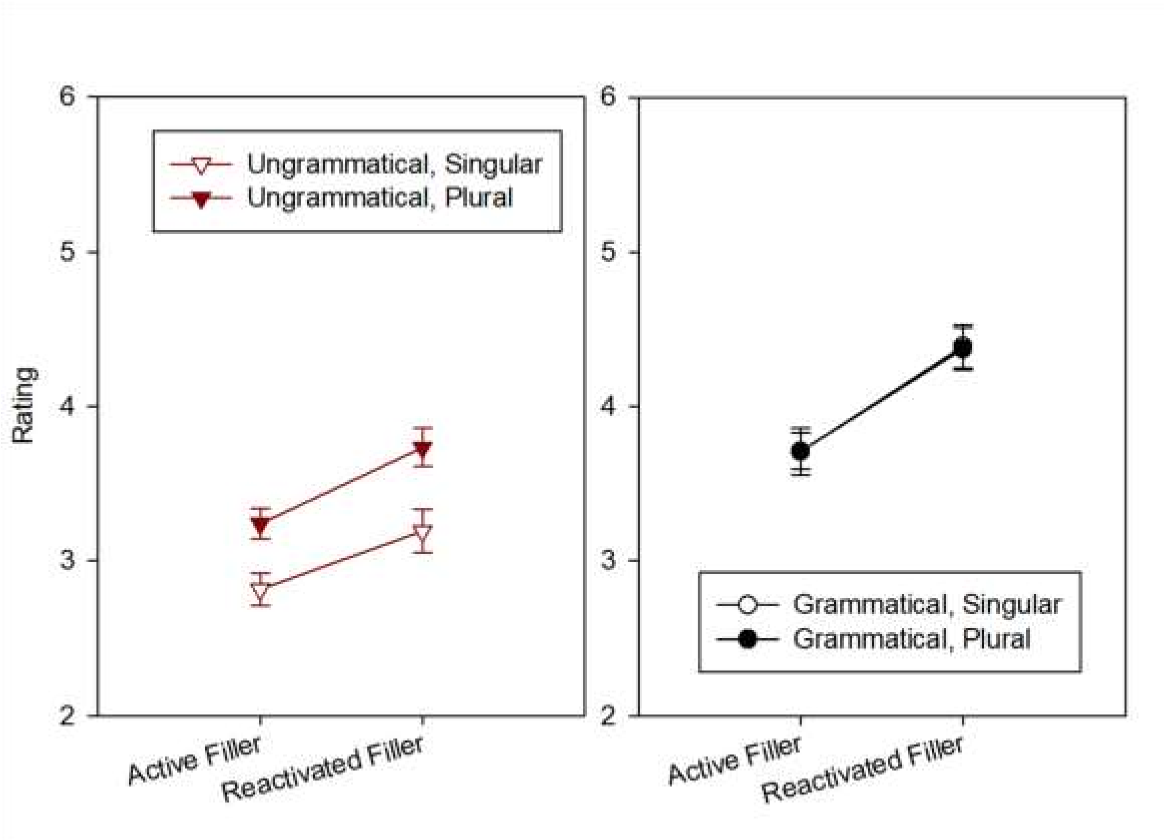


Figure 20. Mean acceptability scores for experiment 5a.

Table 27. Summary of fixed effects from the ordinal mixed effect model in experiment 5a.

Random intercepts were included for subjects and items, as were by-subject intercepts for *Local noun*, and *Grammaticality*, and by-item intercepts for *Local noun* and *Grammaticality*.

	Estimate	SE	z	p
(Intercept)				
Local noun	0.37	0.15	2.47	<0.05 *
Grammaticality	-0.40	0.30	-4.65	< 0.001***
Dependency type	-0.87	0.07	-11.84	< 0.001***
Local noun x *Grammaticality*	0.72	0.14	5.02	< 0.001***
Grammaticality x *Dependency type*	0.42	0.14	2.91	< 0.01**
Local noun x *Dependency type*	-0.01	0.14	-0.07	0.95
Local noun x *Grammaticality* x *Dependency type*	-0.38	0.29	-1.34	0.18

Main effects of all three factors were observed. We observed a main effect of *Local noun* where items with singular local nouns were rated lower than those containing plural local nouns. We observed a main effect of *Grammaticality* where ungrammatical items were rated significantly less acceptable than those containing grammatical ones. Finally, a main effect of *Dependency type* was observed such that items with active Fillers were rated significantly less acceptable than those containing the reactivated Fillers.

We found an interaction between *Local noun* and *Grammaticality* where items containing singular local nouns were rated less acceptable than those containing plural local nouns in ungrammatical conditions, but the same in grammatical conditions. This was further supported by a main effect of *Local noun* ($\beta = 0.67$, $SE = 0.10$, $z = 6.50$, $p < 0.001$) and a main effect of *Dependency type* ($\beta = -0.63$, $SE = 0.10$, $z = -6.10$, $p < 0.001$) in ungrammatical but not grammatical conditions. This indicates an illusion of grammaticality effect consistent with agreement attraction.

However, an interaction between *Dependency type* and *Grammaticality* was also observed such that the differences between the active filler and reactivated filler were larger in grammatical sentences ($\beta = -1.21$, $SE = 0.31$, $z = -3.98$, $p < 0.001$) than in ungrammatical sentences ($\beta = -0.54$, $SE = 0.25$, $z = -2.14$, $p < 0.05$), suggesting that when considered in light of the grammatical sentence baseline, reactivated filler sentences elicit relatively more agreement attraction, with a reduced difference between the grammatical and ungrammatical plural conditions in the active filler (M = 0.47) than the reactivated filler conditions (M= 0.64). Items containing singular local nouns were judged less acceptable than those containing plural local nouns in the Reactivated Filler condition ($\beta = 0.36$, $SE = 0.16$, $z = 2.18$, $p < 0.05$), as well as, marginally, in the Active Filler condition ($\beta = 0.34$, $SE =$

0.19, z= 1.74, p=0.08). Finally, the three-way interaction between *Local noun*, *Filler type*, and *Grammaticality* did not reach significance.

In combination, these results show evidence for attraction in an offline measure for both active and reactivated wh-fillers. The results are consistent with the idea that the difference between the two types of filler was stronger in ungrammatical than grammatical conditions.

2.21. Experiment 5b: A Self-Paced Reading Experiment

2.21.1. Participants, Materials and Design

Participants were 76 native speakers of English from Northwestern University with no history of language disorders. All participants provided informed consent and received credit (1 credit/ 45 minutes) in an introductory Linguistics class.

Critical items were similar to Experiment 3a. Items were distributed in a pseudo-randomized manner to make sure that participants did not receive the same type of experimental itemssequentially. The experimental items were combined with 64 filler sentences of similar complexity. The experiment took around 30 minutes to complete.

2.21.2. Procedure

A similar procedure was employed as with Experiment 3b.

2.21.3. Analysis

Factors are as described in Experiment 3a. The analysis was conducted as described in Experiment 3a. The by-region exclusion percentages due to outlier removal were 2.52 %

(verb region), 3.25 % (spillover region 1), and 2.88 % (spillover region 2).

2.21.4. Results & Discussion

Region-by-region reading times for Active Filler conditions are presented in Figure 21, the Reactivated Filler conditions are presented in Figure 22, and interaction plots for spillover regions 1 and 2 are presented in Figure 23 and 24 respectively. Mixed effect model outputs are presented in Table 28. Mean accuracy for critical trial comprehension questions was 84.0%.

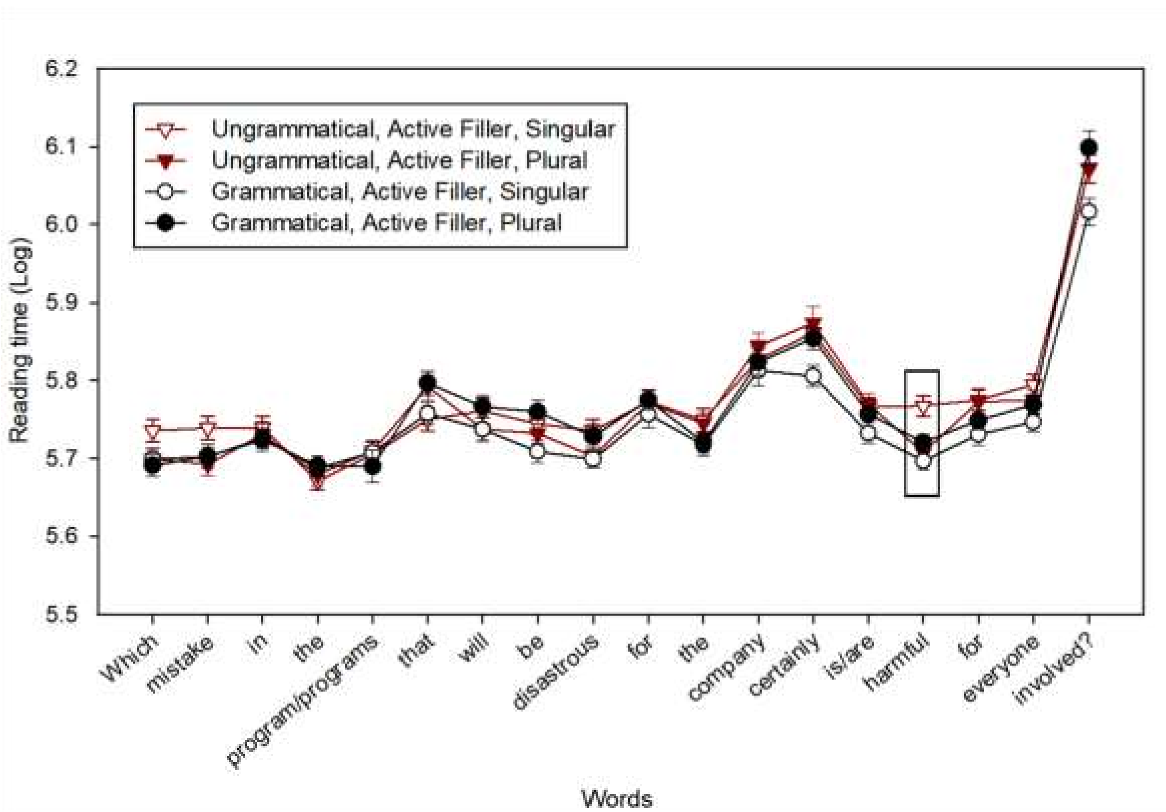


Figure 21. The region-by-region reading times for experiment 5b Active Filler conditions. The box indicates the spillover region 1, *harmful*.

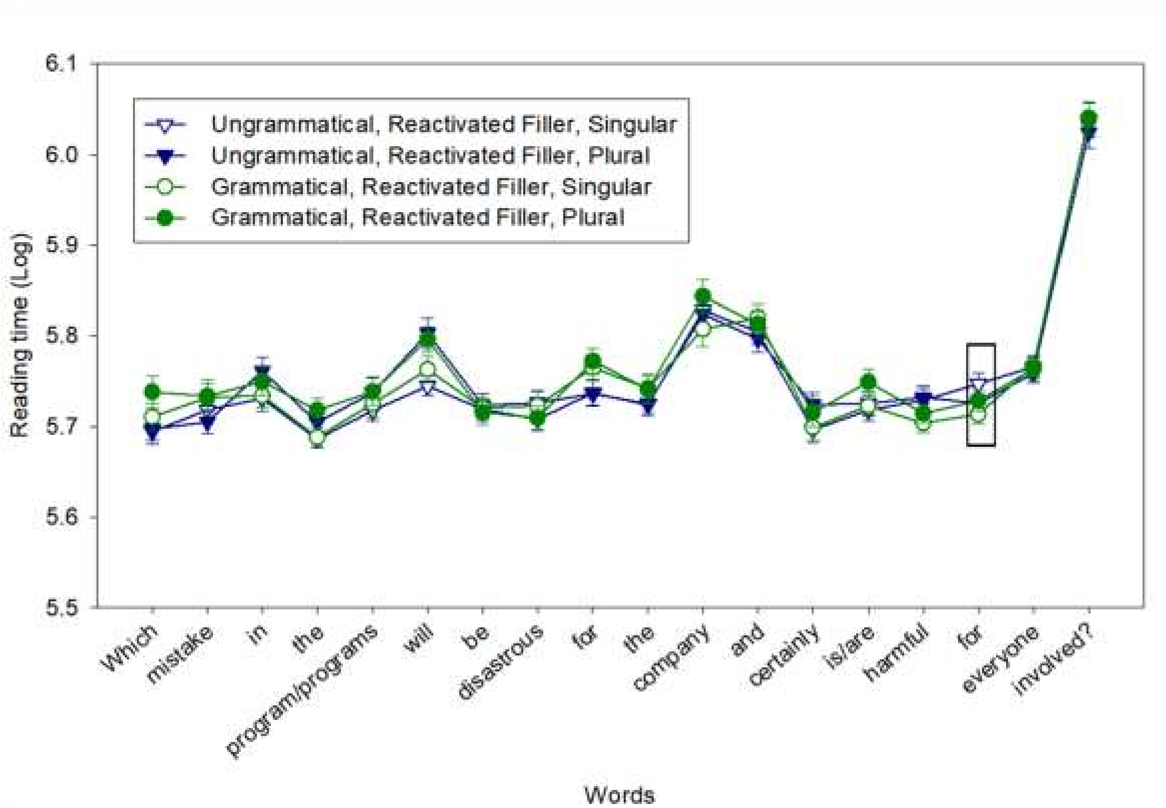


Figure 22. The region-by-region reading times for experiment 5b Reactivated Filler conditions. The box indicates the spillover region 2, *for*.

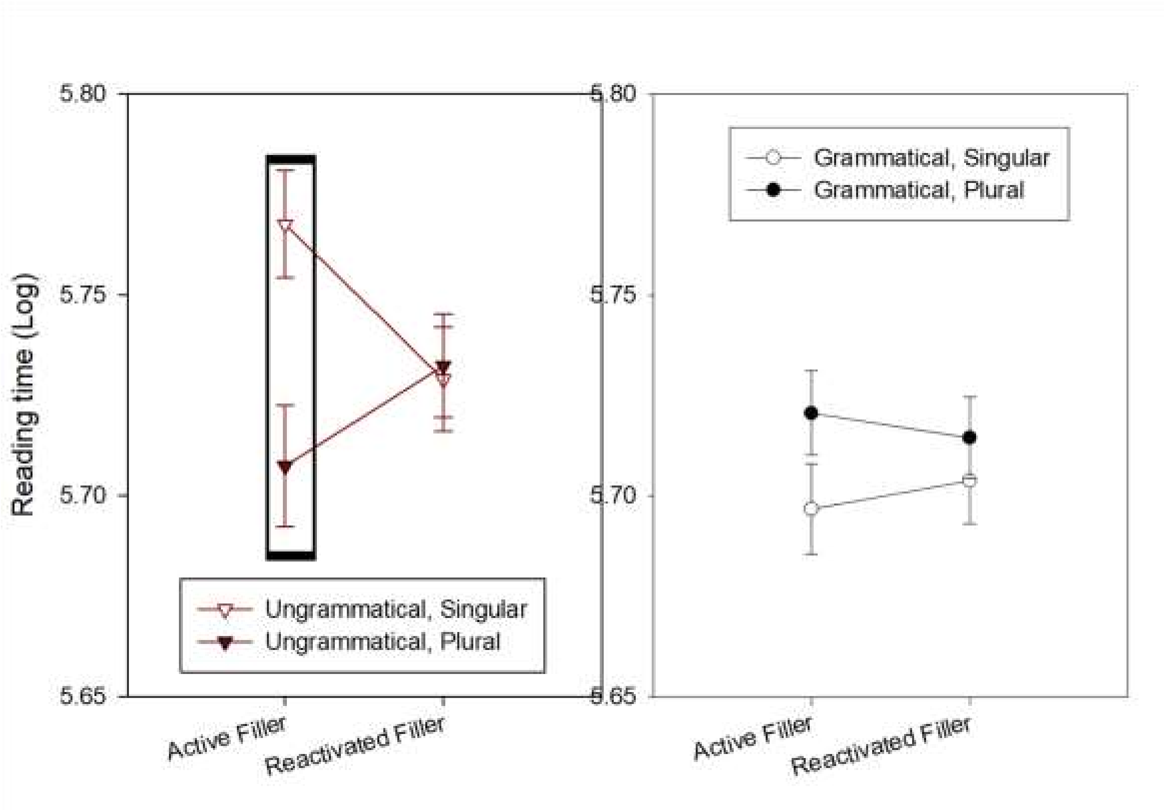


Figure 23. Interaction plot for spillover region 1 (*harmful*).

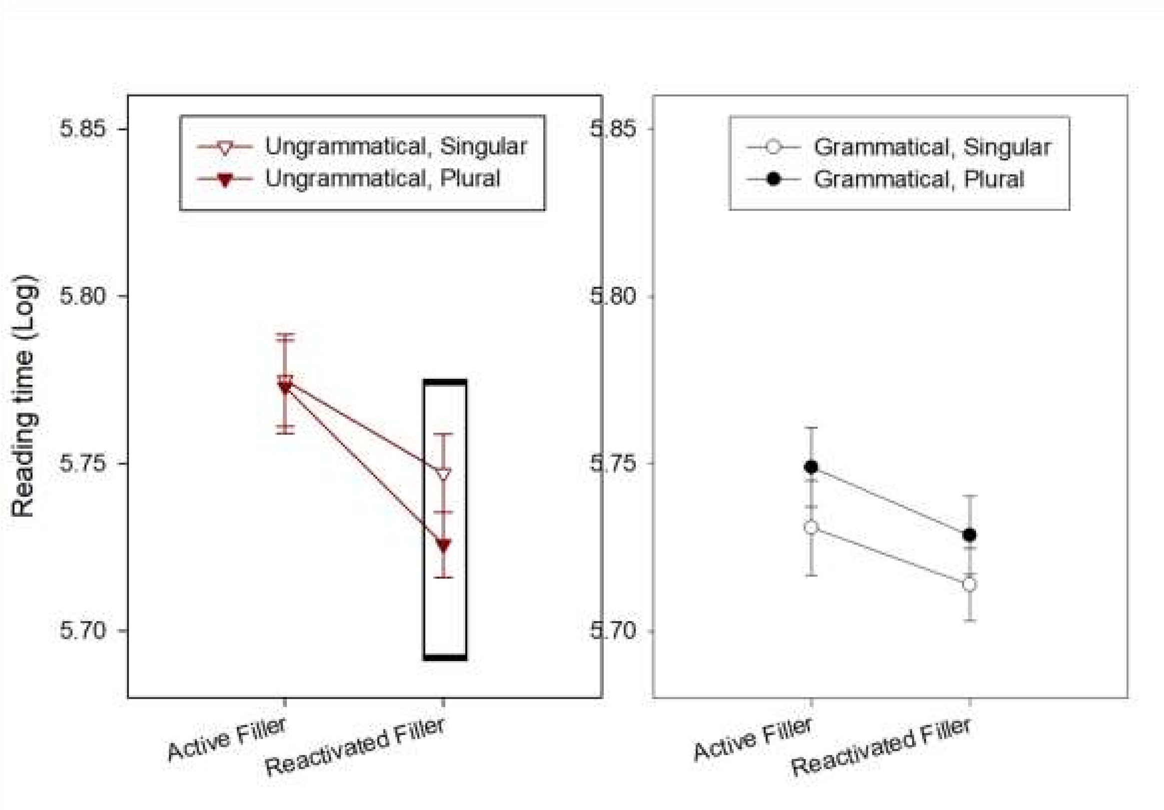


Figure 24. Interaction plot for spillover region 2 (*for*).

Table 28. Summary of results of linear mixed effects models by region in experiment 5b.

	Estimate	SE	t	p
Verb Region (*is/are*): by-subject random intercepts and slopes for *Grammaticality* and *Dependency type*, by-item random intercepts and slopes for *Dependency type*				
(Intercept)	5.74	0.03	214.49	
Local noun	0.02	0.01	1.81	0.07
Grammaticality	0.00	0.01	0.41	0.68
Dependency type	0.03	0.01	2.38	< 0.05 *
Local noun x *Grammaticality*	-0.02	0.02	-1.31	0.19
Grammaticality x *Dependency type*	0.04	0.02	2.18	<0.05 *
Local noun x *Dependency type*	-0.01	0.02	-0.31	0.76
Local noun x *Grammaticality* x *Dependency type*	0.00	0.03	0.15	0.88
Verb Spill-over Region 1 (*harmful*): by-item random intercepts and slopes for *Dependency type*				
(Intercept)	5.72	0.02	231.31	
Local noun	-0.00	0.01	-0.26	0.79
Grammaticality	0.03	0.01	3.49	< 0.001***

Dependency type	0.01	0.01	0.66	0.51
Local noun x *Grammaticality*	-0.04	0.02	-2.54	<0.05*
Grammaticality x *Dependency type*	0.01	0.02	0.70	0.49
Local noun x *Dependency type*	-0.02	0.02	-1.39	0.17
Local noun x *Grammaticality* x *Dependency type*	-0.07	0.03	-2.02	<0.05 *
Verb Spill-over Region 2 (*for*): by-subject random intercepts and slopes for *Local noun*, *Dependency type* and *Grammaticality*, by-item random intercepts and slopes for *Local noun* and *Dependency type*				
(Intercept)	5.74	0.02	236.94	
Local noun	0.00	0.01	0.29	0.77
Grammaticality	0.02	0.01	2.40	< 0.05*
Dependency type	0.03	0.01	2.87	< 0.01**
Local noun x *Grammaticality*	-0.04	0.02	-2.24	<0.05*
Grammaticality x *Dependency type*	0.02	0.02	1.41	0.16
Local noun x *Dependency type*	-0.00	0.02	-0.06	0.95
Local noun x *Grammaticality* x *Dependency type*	0.02	0.03	0.60	0.55

At the verb region, a main effect of *Dependency type* was observed such that items with the active filler were read significantly slower than those containing the reactivated filler.

At the spillover region 1, we observed a main effect of *Grammaticality* where ungrammatical sentences were read significantly slower than their grammatical counterparts. This was qualified by an interaction between *Grammaticality* and *Local noun*, and an interaction between *Grammaticality, Local noun*, and *Dependency type.* Further subset analysis suggest that these differences were driven by the active filler dependency condition, which showed a significant main effect of *Grammaticality* ($\beta = 0.04, SE= 0.01, t= 2.70, p<0.05$) and an interaction between *Local noun* and *Grammaticality* ($\beta = -0.08, SE= 0.02, t= -3.18, p<0.01$). In contrast, for the reactivated filler, there was a marginal main effect of *Grammaticality* ($\beta = 0.02, SE= 0.01, t= 1.84$) but no significant interaction between *Local noun* and *Grammaticality* ($\beta = -0.01, SE= 0.02, t= -0.41, p>0.05$). This indicates more agreement attraction for active versus reactivated wh-fillers at the spillover region 1. Importantly, we found an interaction between *Local noun* and *Dependency type* in the ungrammatical conditions ($\beta = -0.06, SE= 0.03, t= -2.21, p<0.05$) but not in grammatical conditions ($\beta = 0.01, SE= 0.02, t= 0.46, p>0.05$).

At the spillover region 2, we found the critical interaction between *Local noun* and *Grammaticality* where the differences between plural and singular local nouns were larger in the ungrammatical conditions than in the grammatical conditions. We also report the reading times at the adverb and at the verb. At the adverb and the verb, there was a main effect of *Dependency type*, such that active fillers were read significantly slower than the reactivated fillers (Adverb: $\beta = 0.14, SE= 0.02, t=$ 8.61; Verb: $\beta = 0.03, SE= 0.01, t=$ 2.38).

2.21.5 Discussion

The current experiment addresses the question of differences between wh-fillers that are linked to the gap in the matrix clause verb directly (active filler) versus wh-fillers linked to the gap in the first conjunct and subsequently reactivated in the coordinate structure. Offline acceptability results show that the interaction between *Local noun* and *Grammaticality* was numerically larger in the reactivated filler conditions, relative to the active filler conditions. We also observed that in reading time measures, agreement attraction was significantly larger for the active filler than the reactivated filler in spillover region 1, as indexed by the three-way interaction in this region; however, both filler types led to attraction in the following region (spillover region 2), with a two-way interaction between *Local noun* and *Grammaticality*. This suggests, although both the reactivated filler and active filler may lead to an agreement attraction effect, the effect was stronger for active fillers and manifested at an earlier stage, than it did for the reactivated fillers.

We have further observed that the second verb and the adverb preceding the second verb were read significantly more slowly in the active filler conditions than in the reactivated filler conditions. We contend that this means that the active filler was maintained in memory. As we have discussed earlier, one of the motivations for the maintenance of the filler is the integration cost effect (Gibson 1998; Grodner & Gibson, 2005; Warren & Gibson, 2002). The observation that the adverb and the verb are read significantly slower in the active filler conditions than the reactivated filler conditions is the following. The active filler caused a larger integration cost because it was maintained in memory for a long distance and it has been observed that the longer dependency gives rise to the more processing cost at the end of the dependency due to the integration cost (Gibson, 1998; Grodner & Gibson, 2005; Warren & Gibson, 2002). The verb region is where the WhFGD is completed. Furthermore, the adverb can clearly signal the presence of the verb and thus the parser can expect that the verb which can terminate WhFGD is upcoming. As a result, as early as the adverb position, the parser can recognize that the WhFGD is being completed, leading to an integration cost at this point. The reactivated filler, on the other hand, was released from memory, and then reactivated and put into maintenance again at the coordinating connective. The distance between the point where the wh-filler was reactivated (coordinating connective, *and*) and the point where the WhFGD is completed (i.e., the second verb position) was short. Therefore, the integration cost should be smaller accordingly.

If we only assume that retrieval plays a role, we would not predict such difference at the second verb position, as both in the active filler and the reactivated filler conditions, the wh-filler should be retrieved at the second verb position, and the distance between the point where the wh-filler is recognized and the second verb where the wh-filler is to be reactivated are basically the same.

As suggested earlier, the differences in the strength and the timing of agreement attraction could be due to whether or not the parser has previously released the wh-filler from maintenance. For the active filler, information associated with the wh-filler is well preserved because the filler has not been released from maintenance and subsequently reactivated. The maintenance of the wh-filler could make available the detailed information of the wh-filler where the parser could access both the head and the modifier, leading to stronger agreement attraction when there is a mismatch between the verb and the head noun but a match between the verb and the local noun. On the other hand, for reactivated fillers, the parser releases the wh-filler from memory and subsequently reactivates the wh-filler by means of the coordinating connective or the recognition of the gap. Therefore, given that the wh-filler is released from maintenance at an early point in the sentence, the released wh-filler is subject to memory decay. We then expect that the wh-filler is less accessible compared to the active filler and thus the information associated with the filler is not accessible for the parser when the second verb is processed. As a result, the structure of the wh-filler, including information about the head noun and the modifier is less accessible, leading to a lower degree of the agreement attraction in the ungrammatical constructions, and a delay in the timing of attraction as the filler is reactivated in processing.

2.22. General Discussion

The first two experiments attempted to uncover the storage component in different kinds of wh-fillers. One of the ways to investigate the storage component is to examine the processing costs involved in maintaining different fillers before the controlling element is encountered. We tested constructions where the dependency lengths between different WhPs and their controlling element differed, with the aim to examine the processing costs associated with the maintenance during the WhFGD formation process. The results revealed high processing costs at the embedded NP position for *who* in the object position, and *how*, but not for *who* in the subject position and *why*; there are more uncompleted dependency formations involved for *object who* and *how* compared to *subject who* and *why*. This suggests that the dependency formation triggers the release of wh-phrases from memory. Different kinds of wh-fillers are released from memory, where grammatical requirements associated with the fillers are once satisfied. That is, *object who* and *how* are linked to V and VP respectively, where their grammatical requirements are satisfied, and hence released from memory. *Subject who* and *why* are linked to TP (the sentence) where their grammatical requirements are satisfied, consequently released from memory.

Our observations are compatible with Gibson (1998)'s SPLT, which posits that when grammatical requirements are satisfied (when the reader is able to predict the controlling element), the processing costs associated with holding particular elements will be reduced. Our results also suggest that different wh-phrases are released from memory once the controlling element is encountered, and this is triggered by the grammatical requirement and the satisfaction of the interpretation associated with the wh-fillers.

In experiment 3,4, and 5, we examined how the wh-filler is maintained and accessed in two WhFGD configurations. These studies argue for a processing architecture that incorporates both maintenance and retrieval components. Our assumption is that if information about the filler is maintained, and less susceptible to decay, it will be accessed easily when the verb is processed (Wagers & Phillips, 2014). On the other hand, if some information from the filler is susceptible to decay because it is released from maintenance, we expect it to be less accessible for the parser. Differences in what is accessible at the verb lead to differences in agreement attraction for different types of wh-fillers.

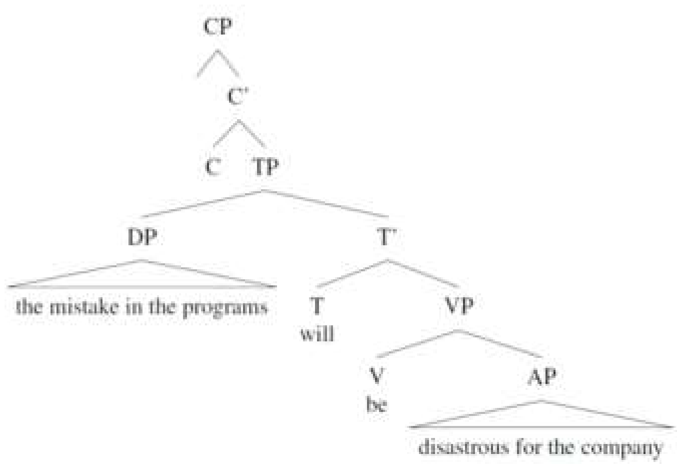
The third experiment tested WhFGD within coordinated structures, in order to examine what information about the wh-filler is accessed at the verb region. According to Wagers & Phillips (2014), information about the category of the wh-filler is maintained throughout the dependency formation process, but thematic and semantic information is not. We investigated whether only category information is maintained, or if details about the content of NP are released from maintenance.

Within the coordinated structure, the wh-filler can be linked to the gap in the first conjunct, and can thus be released from memory. However, the wh-filler should be reactivated when the coordinating connective *and* is processed, due to the CSC and ATB restriction. The results showed that the verb was read faster in the ungrammatical plural local noun conditions than the ungrammatical singular local noun conditions, i.e., we observed an illusion of grammaticality effect. Thus, detailed information associated with the filler (i.e. grammatical information) is readily accessed at the verb, for reactivated wh-fillers.

In the fourth experiment, we compared definite subject NPs with reactivated wh-fillers, in order to understand what motivates the maintenance of an element. In a coordinated structure involving a definite NP in the subject position, the presence of a filler-gap dependency is not signaled, and thus the coordinating connective does not initiate the parser to form a filler-gap dependency in the second conjunct. Thus, until the gap in the second conjunct is encountered, the parser should not construct the structure that involves the filler-gap dependency. Only by recognizing the gap in the second conjunct does the parser register that the definite NP is part of a filler-gap dependency. Thus in this configuration, the parser should not initially register that a filler-gap dependency is involved, and therefore the parser needs to reanalyze the structure as such.

Let us look at the time-course of the resolution of the definite NP versus wh-filler in the WhFGD. For the definite NP, the parser first builds the structure of *the mistake will be disastrous for the company*.

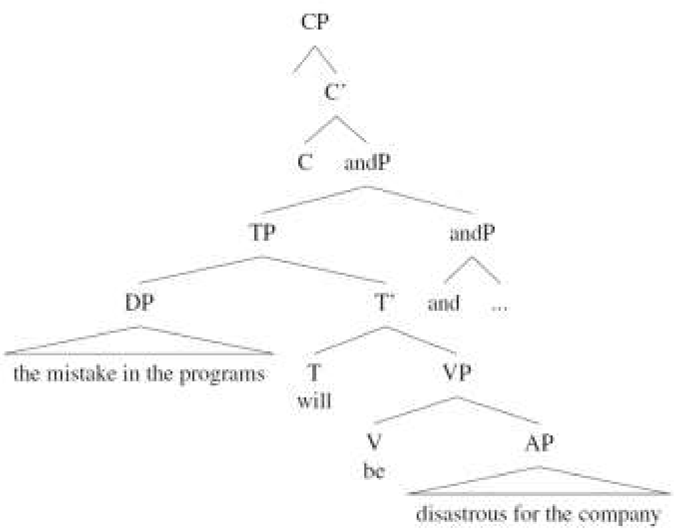
(26) a.



When the parser encounters *and*, it projects andP[16] and connects to the TP.

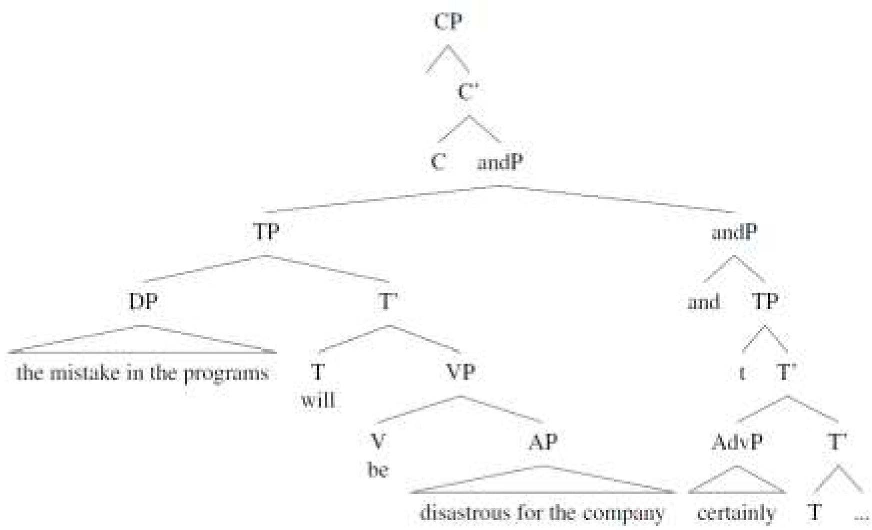
[16] We follow Munn (1993)'s syntax of coordinated structures.

(27)



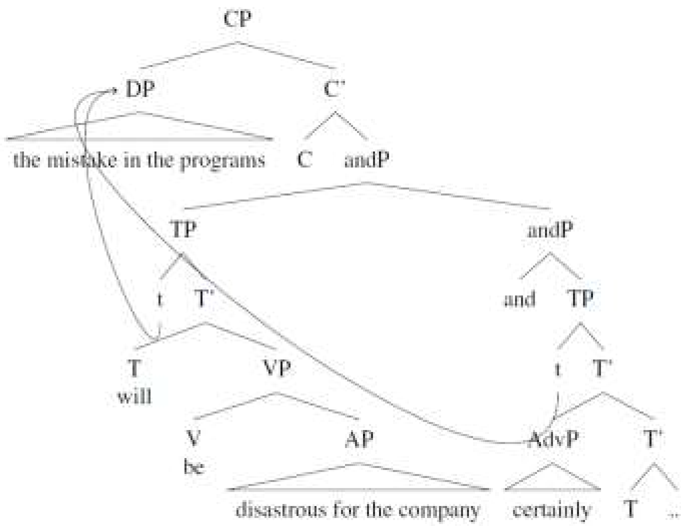
When the parser hits an adverb, *certainly*, the parser undergoes reanalysis process from non-movement to the movement structure. This is so because, in the definite NP conditions, no grammatical constraint signals the presence of the gap in the upcoming conjunct, when the parser encounters the coordinating connective *and*. Unlike the wh-filler, no feature on the definite NP suggests that it is moved, and thus even if the parser encounters the coordinating connective *and*, it cannot recognize that the coordination construction involves movement and is not constrained by CSC and the ATB movement restriction. In other words, the definite NP does not involve a wh-element, and the coordinating connective *and* does not provide cue for the upcoming WhFGD. Upon encountering the coordinating connective *and*, the parser would naturally anticipate a subject followed by a verb driven by the local attachment bias. If this is the case, the parser needs to rebuild the structure that involves a subject gap. Only when the presence of the gap in the second conjunct is encountered, the parser has to reanalyze it into the movement structure that involves FGD. In this case, the parser does not form a WhFGD that applies to the ATB/CSC restrictions. Information associated with the definite NP will not be activated at the coordinating *and*.

(28)



After the sequence of *and certainly*, the parser recognizes the gap in the second conjunct, and only at this time, the parser posits that there is a filler-gap dependency. This means that the parser needs to reanalyze the non-movement structure to the movement structure after realizing the presence of the gap.

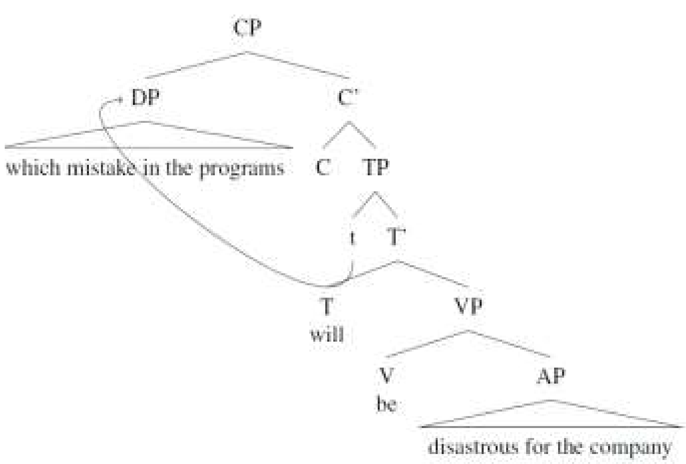
(29)



Therefore, the non-movement to movement structure accounts for the observed reanalysis costs for the definite NP.

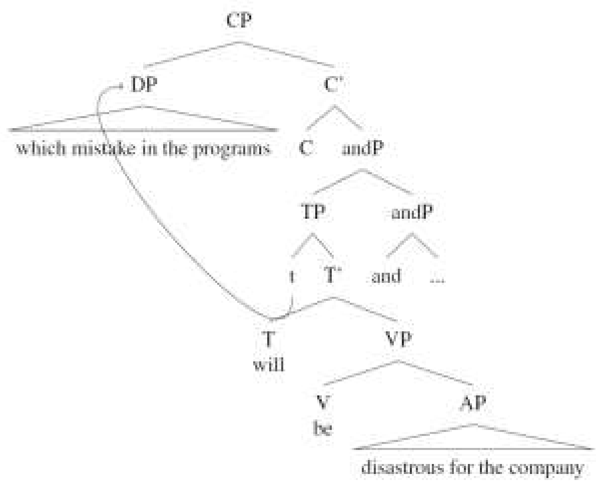
For the case of reactivated filler, the parser first builds the structure of *which mistake will be disastrous for the company*.

(30)



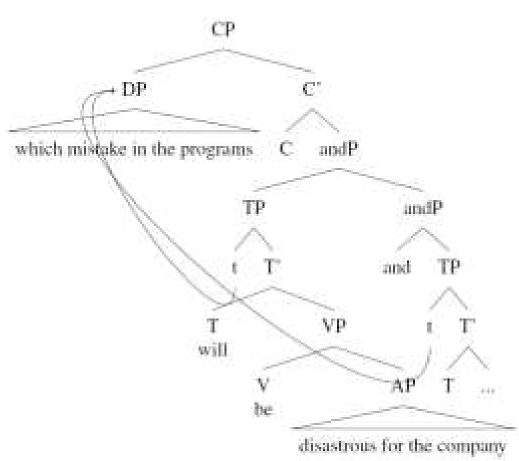
When the parser encounters *and*, it projects andP and connects to the whole sentence.

(31)



When the parser hits the coordinating connective *and*, the parser forms another WhFGD where it actively links the wh-filler to the subsequent gap. The parser will retrieve the wh-filler (*which mistake in the programs*) upon encountering the coordinating connective *and* due to CSC and the ATB restriction.

(32)



As the filler-gap dependency is not constructed initially, the definite NP should not be maintained in memory and, therefore, should be subject to memory decay. The results of Experiment 4 are compatible with the view that a wh-element is different from a definite NP with respect to retrieval; attraction effects are obtained at the main verb, supporting the idea that memory encoding of the filler includes richer information than just its category information (c.f., Wagers & Phillips, 2014, but see also Chow & Zhou, 2019).[17] A numerically weaker illusion of grammaticality effect was observed in the definite NP conditions than in the wh-filler conditions, which follows from the premise that the definite NP is not maintained in memory and is not reactivated upon encountering the coordinative connective *and.* However, we did not find a three-way interaction between *Local noun*, *Grammaticality*, and *Filler type*, which is expected if the accessibility to the wh-filler is different. This means that the magnitude of the attraction effect did not significantly differ depending on the type of the dependency, but it only differed numerically. Also, note that there were differences between the online and offline experiments in that the results of the acceptability rating experiment revealed an agreement attraction both in wh-filler and the definite NP. This could have been due to the possibility for readers to look back to the prior context in the acceptability rating experiment, and may also have been due to the lack of power/more noise in the reading experiment.

The last experiment examined how active fillers and reactivated fillers differed in terms of their maintenance, comparing the accessibility of the wh-filler in these two

---

[17] We thank an anonymous reviewer in *Language, Cogntion, and Neuroscience* for pointing this out.

constructions, when the verb is processed. Both the active filler and the reactivated filler showed an illusion of grammaticality effect, with the reactivated filler eliciting attraction later in processing, as revealed by the three-way interaction of *Local noun*, *Grammaticality*, and *Dependency type* at the spillover region 1 and the interaction between *Local noun* and *Grammaticality* at the spillover region 2. Importantly, at the spillover region 1, we found an interaction between *Local noun* and *Dependency type* in ungrammatical sentences. We contend that this is because in the reactivated filler construction with the coordination structure, the wh-filler is released from memory and the parser reactivates the wh-filler at the point of the coordinating connective, or at the second verb position. The unmaintained information is thus subject to memory decay. In contrast, information associated with the active filler is likely to be maintained, because there is no gap which can complete the WhFGD prior to the gap in the matrix clause, i.e., the active filler is not released from memory and, thus, not subject to memory decay.

We argue that both maintenance and retrieval play crucial roles in the resolution of WhFGD, and adopting either the retrieval or the maintenance view cannot account for the data (see Wagers & Phillips, 2014 for related discussion). To understand this point, let us first assume that only retrieval plays a role in dependency resolution (Nicenboim et al., 2015). In this case, for both the active and reactivated filler, the wh-filler is expected to be reactivated at the same point in the sentence. In the case of the active filler (*Which mistake in the program/s that will be disastrous for the company certainly is/are harmful for everyone involved?*), the recognition at the matrix verb and the recognition of the gap in the second conjunct at the same time triggers retrieval of the wh-filler. In the case of the reactivated filler (*Which mistake in the program/s will be disastrous for the company and certainly is/are harmful for everyone involved?*), the coordinating connective *and* triggers the reactivation of the wh-filler due to CSC and ATB restrictions or the second verb triggers the reactivation of the wh-filler. Thus, in terms of retrieval, we do not expect any difference between the active filler and the reactivated filler. In the cue-based retrieval model, this reactivation prior to the gap should increase activation specifically for the head of the wh-phrase, not the modifier. This would mean that less attraction should be predicted for the reactivated filler, relative to active filler.[18] However, the results show that the active filler reveals agreement attraction at an earlier stage. Thus, the differences between the active and the reactivated filler suggest a role for maintenance in parsing, as information about the active filler should be maintained relatively well whereas that of the reactivated filler should not. The earlier agreement attraction for the active filler suggests that details about the content of the NP is not released from the maintenance.

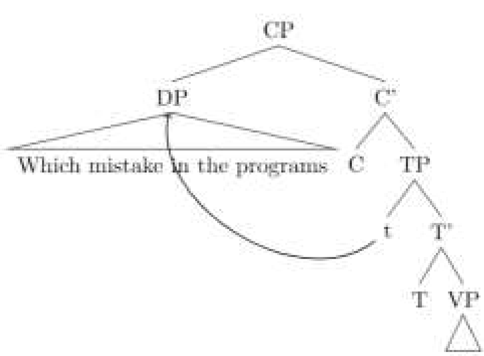
These results suggest that both the maintenance and retrieval are involved in the online WhFGD formation. We showed that category information and the internal structure associated with the filler are accessed at the verb position. However, differences in the accessibility of information with respect to different types of fillers and dependency types cannot be explained if we only posit that information is retrieved from the content-addressable memory store based on the cue-based retrieval mechanism. Our results support

[18] Note that, under the cue-based retrieval model, if the gap increases activation specifically for the head of the wh-phrase but not the local noun (Nicenboim et al., 2015), then the weaker agreement attraction is predicted for the reactivated filler relative to the active filler. We would like to note this as a possible alternative hypothesis.

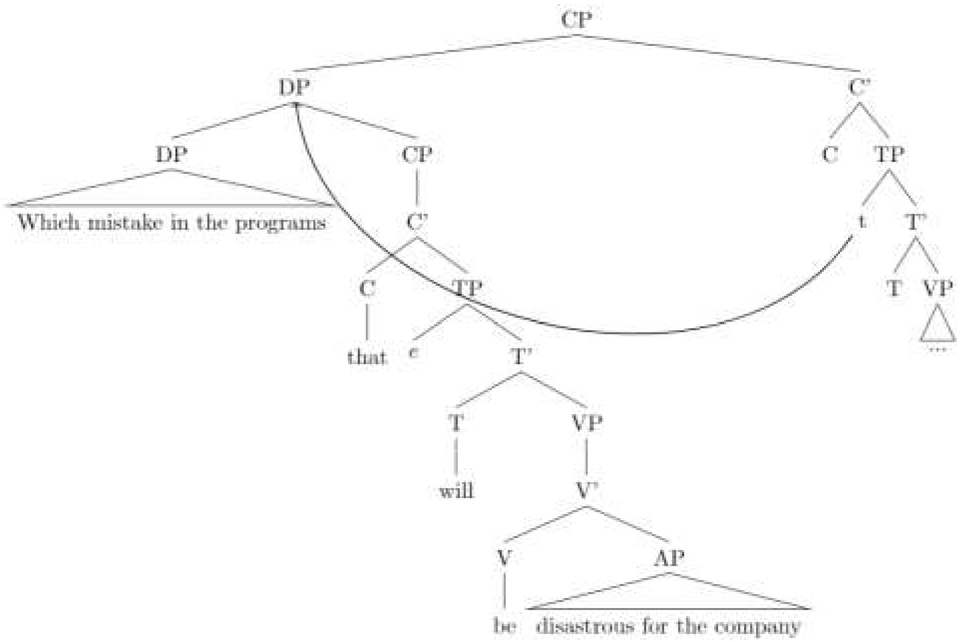
that some information associated with the filler is easily accessed at the verb, whereas some information associated with the filler is hard to access. Namely, the differences in the retrieved information between different fillers (wh-filler vs. the definite NP) and dependency types (reactivated filler vs. active filler) could be attributed to differences in maintenance. If we assume that maintained information leads to greater accessibility of information to the parser, we could account for the differences in the retrieved information between different filler types and dependency types. We showed that both of these two components are used for online WhFGD formation process, where detailed information associated with the filler can be maintained in memory, making it less susceptible to memory decay. On the other hand, if the filler is not maintained in memory, detailed information can still be retrieved at a later stage, though it is subject to decay.

In terms of the time-course of the resolution of the active filler in the WhFGD, even though the wh-phrase, *which mistake in the programs* is modified by the relative clause (*that will be disastrous for the company*), the wh-verb dependency is only established at the matrix verb (*is/are*).

(33)



(34)



The crucial differences between these two dependency types are that for the reactivated filler, the wh-phrase is linked to the gap in the first conjunct and thus the wh-dependency is completed once the parser reaches the first conjunct. This means that the wh-filler can be released from memory when the wh-gap dependency is formed in the first conjunct. When the coordinating connective *and* is encountered, the wh-phrase needs to be reactivated, but the wh-filler was released from the maintenance before *and* is processed. Unmaintained information is subject to memory decay, thus before *and* is processed, the reactivated filler has been subject to decay. *And* indeed reactivates the filler, but some information associated with the filler could have decayed due to the release from the maintenance. Therefore, detailed information associated with the wh-filler could not be accessed at the verb position. Conversely, the active filler must be linked to the gap in the matrix clause directly in the second conjunct. This may allow for stronger maintenance for detailed information associated with the wh-phrase such as category information and the internal structure of the wh-phrase. Thus, to account for the difference between active and reactivated fillers, we need to consider both maintenance and retrieval mechanisms.

3. Retrieval[19]

3.1. Introduction of Syntax of NPE Puzzle

Part of the study of ellipsis is concerned with structure within the ellipsis site (Lasnik, 2001, 2005; Merchant 2001, 2005, 2013). There are roughly two possibilities: either the omission site is associated with certain syntactic structures (*Surface Anaphora*: Hankamer & Sag, 1976; Sag & Hankamer, 1984) or it is associated with certain pro-forms (*Deep Anaphora*: Culicover & Jackendoff, 2005; Hankamer & Sag, 1976).

NP-Ellipsis (henceforth NPE) in English seems to show deep anaphora properties. NPE in (35a), like Anaphoric *one* in (35b), does not tolerate wh-extraction out of the ellipsis site (the NPE-site), even though the same NP in a non-ellipsis context allows wh-movement as in (35c) (Davies & Dubinsky, 2003). Note that in (35a), demonstratives are contrasted and thus the largest deletable constituent should be NP *[$_{NP}$ stories about]* (Fox & Lasnik, 2003; Merchant, 2008; Takahashi & Fox, 2005).[20]

---

[19] Portions of the Chapter 3 have been published in *Language, Cognition, and Neuroscience* (Kim, N., Brehm, L., & Yoshida, M. (2019). The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval. *Language, Cognition and Neuroscience*, *34*(2), 190-213).

[20] We would like to note that the judgments we are reporting here are relative judgment rather than absolute judgment. Even though the judgment of some cases are subtle, native speakers who we interviewed reported the contrasts we show in this study.

(35) a. *Mary can tell who he wrote THESE stories about t_$_{who}$, but she cannot tell who he wrote THOSE [$_{NP}$ ~~stories about t~~ _$_{who}$].

b. *Mary can tell who he wrote these long stories about t_$_{who}$, but she cannot tell who he wrote these short ones.

c. Mary can tell who he wrote THESE stories about but she cannot tell who he wrote THOSE stories about.

The restriction on wh-movement out of the NPE-site follows straightforwardly from the pro-form analysis of NPE (Lobeck, 1995, 2007; Panagiotidis, 2003; Postal, 1969; Ross, 1967). Under this account, because the missing NPE-site is associated with a pro-form which does not have any internal structure, the NPE-site does not involve the position of a wh-trace (Chisholm, 2003; Merchant, 2013a, 2014).

NPE contrasts sharply with TP-Ellipsis (TPE) and VP-Ellipsis (VPE) in this respect. Both TPE and VPE tolerate wh-extraction out of the ellipsis site, suggesting that the ellipsis sites in these constructions have internal structures that can contain the trace of the wh-moved elements.

(36) a. John knows which professor we invited, but he is not allowed to reveal which one ~~[$_{TP}$ we invited t_$_{which\ one}$]~~.

b. Mary doesn't know who we can invite, but she can tell you who we can NOT ~~[$_{VP}$ invite t_$_{who}$]~~. (Takahashi & Fox, 2005:4)

With this background, we show that NPE in English involves ellipsis, i.e., that the NPE-site indeed has an internal syntactic structure. We further show that the restrictions on wh-extraction from the NPE-site follow from independently motivated requirements. Uncovering whether NPE involves an internal structure or not will inform us of the retrieval mechanisms that we aim to investigate later.

3.2. Is NP-Ellipsis Ellipsis or Pro-form?

Extraction is one of the strongest arguments for internal structure in the ellipsis site (Merchant, 2013a:538). However, as we have seen, extraction out of the NPE-site is not possible.[21] Besides extraction, how can we show that the NPE-site is associated with silent syntactic structures? We would like to show two new arguments. The first argument is based on Binding Condition C connectivity effects (Chomsky, 1981). To see those effects, we look at the NPE construction, where the PP object escapes the NPE-site, so-called Nominal Gapping (NG) as in (37) (Yoshida, Wang, & Potter, 2012).

(37) John's book of music from Blackwell and Mary's ~~book t_{PP} from Blackwell~~ [$_{PP}$ of

---

[21] There are few diagnoses for ellipsis such as missing antecedent (Chisholm, 2001; Grinder & Postal, 1971), sloppy identity (Hardt, 1991; Johnson, 2001; Merchant, 2013a; Ross, 1967; Sag, 1976), and vehicle change (Fiengo & May, 1994; Merchant, 2001:24). However, as Merchant (2013a) convincingly shows that these effects are seen in non-ellipsis contexts as well (Merchant, 2013a: 539-541), these diagnoses are thus not strict diagnostics for ellipsis. Therefore, following Merchant (2013a), we do not use them as the tests for ellipsis.

philosophy] are both highly recommended.

NG is a variant of NPE as the licensing condition of the NPE and NG are the same. For example, both NG and NPE can co-occur with a possessor NP like *Mary's* but not with an attributive adjective like *new* as in (38) (for more details, see Yoshida et al., 2012:14).

(38) *I read John's old book of music and Mary's (*new) [$_{NPE}$ ~~book~~] (of philosophy).

In a non-elliptical DP context like (39), a pronoun in the direct object position and the name in the indirect object cannot be coreferential, the Binding Condition C effect (Chomsky, 1981), i.e., the pronoun c-commands the name as in (39a). However, when the pronoun is embedded within a larger NP and c-command relation is removed, coreference is possible as in (39b). This anti-coreference effect cannot be attributed to the presence of a pronoun within the adverbial clause (the *although*-clause) as (39c) is acceptable in the coreferential reading.

(39) a. *Although John's introduction of Beth to David was polite, [$_{TP}$[$_{DP}$ Bill's introduction [of **her**$_1$] [to **Sue**$_1$'s future colleagues]] was not].

b. Although John's introduction of Beth to David was polite, [$_{TP}$[$_{DP}$ Bill's introduction [of **her**$_1$ future colleagues] [to **Sue**$_1$]] was not].

c. Although John's introduction of **her**$_1$ to David was polite, [$_{TP}$[$_{DP}$ Bill's introduction of Joe to **Sue**$_1$'s future colleagues] was not].

If we see the contrast similar to (39a-b) in the NPE context, we can argue that the NPE-site involves the same structure as (39). To show this, we have conducted an acceptability rating experiment (Mahowald, Graff, Hartman, & Gibson, 2016; Sprouse & Almeida, 2017). In this experiment, the participants were asked to rate the naturalness of a sentence in which two NPs in the sentence are referring to the same person (Gordon & Hendrick, 1997; Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007; Yoshida, Potter, & Hunter, 2018). To strengthen our argument, we compared Binding Condition C (BCC) effects in the NPE context against those in an Anaphoric *one* context, which is arguably deep anaphora and thus ellipsis is not involved. Anaphoric Element (NPE vs. Anaphoric *one*) x Pronoun Positions (the pronoun embedded within a DP vs. the pronoun not embedded within a DP) were manipulated in a 2×2 factorial design. In each stimulus sentence, participants were asked to rate the naturalness when these two pronouns and nouns referred to the same person. Eight native speakers of English at Northwestern University participated in the experiment. The sample set of stimuli are summarized in the Table 29.

Table 29. Sample stimuli of BCC reconstruction effects.

		Pronoun	
		C-command	*Not C-Command*
Anaphoric element	*NPE*	Although John's introduction of her to David was polite, Bill's to Sue's future colleagues wasn't.	Although John's introduction of her future colleagues to David was polite, Bill's to Sue wasn't.
	One	Although John's long introduction of her to David was polite, Bill's short one to Sue's future colleagues wasn't.	Although John's long introduction of her future colleagues to David was polite, Bill's short one to Sue wasn't.

If the NPE-site involves the structure of NP akin to (39), and if, following Yoshida et al., (2012), the direct object is (right-ward) extracted from the NPE-site, then we expect that the NPE examples should show the BCC reconstruction effect. Thus, when the pronoun c-commands the name as in the NPE/c-command condition, the coreference should be blocked, but if the c-command relation is removed, the coreference should be possible. Anaphoric *one*, which does not have the internal structure, should not show such contrast. The results were analyzed by a Linear Mixed Effect Model (LME), which revealed a significant interaction of the *Anaphoric element* x *Pronoun* ($\beta = -1.93, SE=0.54, t=-3.57, p<0.01$) and no main effect of neither *Anaphoric element* nor *Pronoun*. Pairwise comparisons revealed that NPE/C-command condition was significantly less acceptable than the NPE/Not C-command conditions ($\beta = 1.03, SE=0.18, t=5.74, p<0.001$). There were no such significant differences in the Anaphoric *one* conditions (p>0.1). These results indicate that the BCC reconstruction effect was observed in the NPE context. The contrast between the NPE conditions and the Anaphoric *one* conditions follows straightforwardly if the NPE-site involves silent syntactic structures but Anaphoric *one* does not.

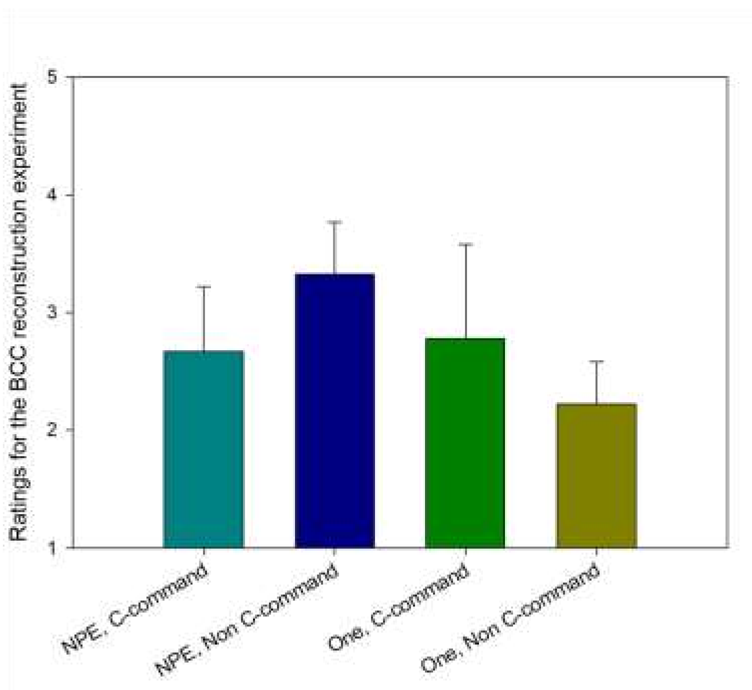


Figure 25. Acceptability ratings of BCC reconstruction effects.

The second argument is based on the scope of the quantifier embedded within the ellipsis site (Collins, 2015, 2018; May, 1977, 1985; Merchant, 2013a). Merchant (2013a:539) shows that a quantifier within an ellipsis site can out-scope elements outside the ellipsis site. However, this inverse scope effect is missing in a corresponding non-elliptical pro-form construction. Merchant (2013a:539) cites the following pair of examples.

(40) a. A doctor examined every patient, and then a nurse did. ($\exists > \forall$, $\forall > \exists$)

b. A doctor examined every patient, and then a nurse did it. ($\exists > \forall$, $*\forall > \exists$)

In (40), the universal quantifier can out-scope the existential quantifier; however, in an example of the *do-it* VP-anaphora, the inverse scope reading is unavailable. The inverse scope reading is available in the non-ellipsis counterpart of (40a), such as (41).

(41) A doctor examined every patient, and then a nurse examined every patient too.

The examples in (40) and (41) suggest that the VPE-site in (40a) contains a universal quantifier in the same way as in (41), but not in VP-anaphora like (40b).

If NPE shows the similar inverse scope reading, we can make an argument for the ellipsis analysis of NPE. To test whether the inverse scope reading is available in the NPE context, we have conducted an acceptability rating experiment[22]. In this experiment, a target

[22] See Collins (2015) for similar arguments in the adjunct PP deletion context.

sentence involving quantificational NPs is followed by sentences that set up the context, like in Table 30[23]. Participants were asked to rate how natural the first sentence is in the context illustrated by the context sentences. The target sentences are compatible with either an inverse scope reading or a surface scope reading for the first sentence. Thus, if, the inverse scope reading is not available in the first sentence, then it should be rated low in the inverse scope context. Anaphoric element (NPE vs. Anaphoric *one*) x Context (Inverse Scope vs. Surface Scope) are manipulated in 2×2 factorial design. Participants were 12 native speakers of English at Northwestern University. A sample set of stimuli are summarized in the table 30.

Table 30. Sample stimuli of inverse scope experiment.

		Context	
		Inverse Scope	*Surface Scope*
Anaphoric element	*NPE*	John investigated a different doctor's examination of every patient and Mary investigated a different nurse's. Context: Mary is an investigator who looks at nurses' examinations. It seems like the nurse had piles of reports for every patient	John investigated a different doctor's examination of every patient and Mary investigated a different nurse's. Context: Mary is an investigator who looks at nurses' examinations. It seems like the nurse had a report for one particular patient.
	One	John investigated a different doctor's examination of every patient and Mary investigated a different nurse's recent one. Context: Mary is an investigator who looks at nurses' examinations. It seems like the nurse had piles of reports for every patient.	John investigated a different doctor's examination of every patient and Mary investigated a different nurse's recent one. Context: Mary is an investigator who looks at nurses' examinations. It seems like the nurse had a report for one particular patient.

---

[23] Following an anonymous reviewer's suggestion from *Linguistic Inquiry*, we included an adjective *different* in the stimuli. *Different* is licensed by *plural event* (Carlson 1987:545) which is available, in our experiment, when the universal quantifier takes scope over *a different NP* and yields the distributive reading. Thus *a different NP* forces the inverse scope reading, and is not compatible with surface scope context.

The results are analyzed by a Linear Mixed Effect model (LME), which revealed a significant interaction of *Anaphoric element* x *Context* ($\beta = 1.74, SE=0.58, t=2.99, p<0.05$) and no main effect of either *Anaphoric element* or *Context*. Pairwise comparisons revealed that the NPE/Inverse Scope condition was rated significantly higher than the NPE/Surface Scope condition ($\beta = 1.22, SE=0.50, t=2.42, p<0.05$), but no such difference was found in the Anaphoric *one* conditions. This result suggests that, in the NPE context the inverse scope reading is available, but not in the Anaphoric *one* context.

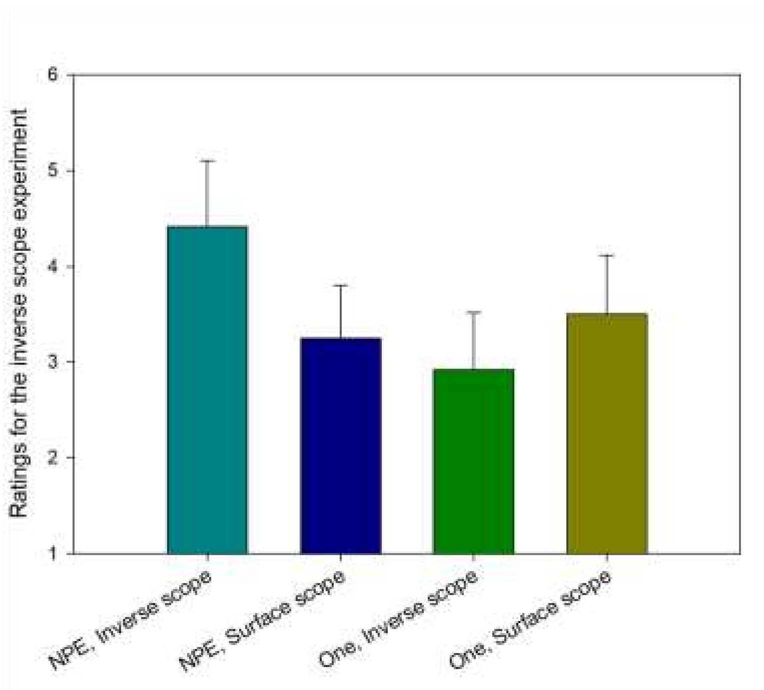


Figure 26. Acceptability ratings of inverse scope experiment.

The difference between the NPE and Anaphoric *one* follows if the NPE-site is associated with syntactic structure that includes the quantificational DP, but Anaphoric *one* is not.

3.3. The Puzzle

Why is wh-movement not possible out of the NPE-site, even though the NPE-site involves silent syntactic structures? Following Aelbrecht (2010, 2016) we argue that wh-movement is blocked since the phase head which attracts wh-elements (Chomsky, 2001) and the ellipsis-licensing head (Aelbrecht, 2010; Lobeck, 2007; Merchant, 2001) are different.

3.3.1. Which Head Licenses NP-Ellipsis?

It has been suggested that NPE can be licensed by multiple heads within DP (Lobeck 1995, 2007). For example, possessor DPs which are assumed to be generated in the specifier of the DP, Plural Demonstratives, which are in the D-head, and Numerals, which are in the Num-head, can co-occur with the NPE-site (Bernstein, 2001; Kester, 1996; Lobeck, 1995).

(42) I read John's book and you read Mary's/those/two ~~[$_{NP}$ book].~~

On the other hand, the definite article *the*, the indefinite article *a(n)*, attributive adjectives, and the singular demonstrative *this/that* cannot be immediately followed by the NPE-site as examples in (43) demonstrate.

(43) a. *I read the book of music and Mary read the/a ~~[$_{NP}$ book of music]~~ too.

b. *I read difficult books and Mary read easy ~~[$_{NP}$ books].~~

c. *I read that linguistics book and he read this ~~[$_{NP}$ linguistics book].~~

Some of the combinations of these DP-internal elements can also be immediately followed by the NPE-site. The possessor DP can be followed by the numeral, and these can be followed by the NPE-site. Similarly, the demonstrative and the numeral and the definite determiner and the numeral can be followed by the NPE-site as well.

(44) I read John's three books and you read Mary's/these/the/two ~~[$_{NP}$ books]~~.

When an attributive adjective is present before the NPE-site, NPE is not licensed.

(45) *I read John's difficult books and you read Mary's/these/the/two easy ~~[$_{NP}$ books]~~.

These observations suggest that the ellipsis licensing head and the NPE-site need to be adjacent. It has been sometimes argued that the ellipsis-site can be long-distance licensed by an ellipsis licensing head if they stand in an Agree relation (Aelbrecht, 2010; Park, 2016). However, this does not seem to be the case in English NPE. If so, we would expect (45) to be possible because there are functional elements that can license the NPE-site, and they can stand in an Agree relation (these elements c-command the NPE-site).

On the other hand, as shown in (45), whether *the* is present or not, when the number element is present, NPE is licensed. Even without the presence of the lexical D head, NPE can be licensed, but merely having the ellipsis-licensing D in the DP is not sufficient to license NPE, as adjectives block NPE. This suggests that D is not an ellipsis licensing head. On the other hand, a numeral can be immediately followed by the NPE-site, whatever D the DP has. This suggests that a functional head related to the numerals in the nominal phrase is the one that licenses NPE (Alexiadou & Gengel, 2012). This conclusion is compatible with Lobeck's (1995:89) observation that NPE is licensed mostly by elements that are specified as plural. We contend that the head related to the numeral is the Number-head (Alexiadou & Gengel, 2012; Bernstein, 2001; Lobeck, 1995, 2007; Ritter, 1991).

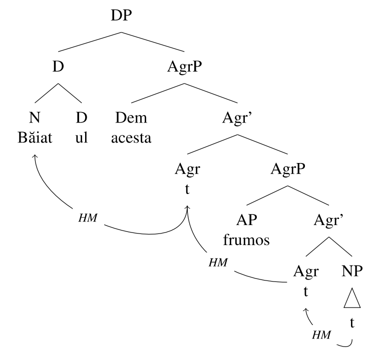
How can the NPE-site be licensed when an overt Number-head is not present in DP (e.g., *I read John's books and you read Mary's*)? It has been suggested that the Number-head or a functional head generated lower than D is responsible for the number marking of NP and thus plural marking of the head-noun is due to the selectional relation between the Number-head and the head noun (Lobeck, 1995:89). Number specification is always implicated, and the NumP or a functional head responsible for number specification is always present within the DP. Furthermore, cross-linguistic considerations suggested that demonstratives and possessive DPs are originally generated in the Spec_NumP or a functional projection below DP, and then moved to the Spec_DP (Alexiadou, Haegeman, & Stavrou, 2007; Brugè 1996, 2002; Brugè & Giusti, 1996; Giusti, 1997, 2002; Guardiano, 2010; Panagiotidis, 2000; Roberts, 2011; Shlonsky, 2004). For example, Giusti (2002) cites the following example from Romanian, showing that in some languages, demonstratives can be generated lower than D. For example, in (46), the N-head *boy* undergoes head movement to D, moving across the demonstrative, *acest, this*. Assuming the Head Movement Constraint (HMC; Baker 1988; Travis 1984), the demonstrative element in (46) cannot be a head, because if it is, N must skip the intermediate demonstrative head position and violate HMC. Giusti argues that demonstratives are a phrase, not a head, generated in a specifier of the functional projection which is generated lower than D.

(46) a. Băiatul acesta frumos

Boy-the this nice d to the Spec_DP

'this nice boy'. (Giusti, 1997:107)

b.



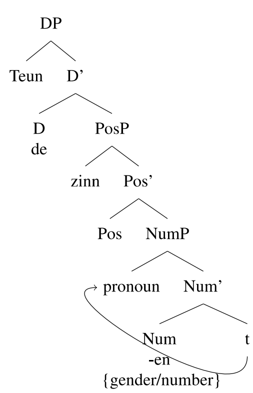
English demonstratives inflect for number, and it follows naturally if demonstratives are generated within the NumP and stand in a Spec-Head agreement relation with the Num-head. Similarly, in some languages, the possessor DP is generated in the NumP. Corver and van Koppen (2010) cite examples of Posessor Doubling in Asten Dutch as shown in (47).

(47) a. ik vein Teun de zinn-en echt geweldig

I find Teun the his-GEN really great

'I find Teun's really great'. (Corver & van Koppen, 2010:131)

b.



An important point about the example in (47) for us is that the possessor pronoun is generated lower than DP. Cover & van Koppen (2010) argue that there is a pronoun that engages Spec-Head agreement with Num^0 and inflects for gender and number that is specified by Num^0. This example suggests that the possessor can be generated in NumP. Adopting these analyses, we can plausibly assume that possessive DP and demonstratives, which apparently license NPE, are not heads, but phrases generated lower within the DP and moved to the Spec_DP position. If this analysis is correct, then when NPE is apparently licensed by possessives or demonstratives, the NumP that hosts these elements is present and the Num-head is adjacent to the elided NP.

There is one further piece of supporting evidence for this position. If the demonstrative is in the D-head position, and if demonstratives can license ellipsis, then the demonstrative should be able to license NumP-ellipsis[24]. In this light, (48), where the numeral element is absent in the second conjunct, can be analyzed as NumP-ellipsis. However, the interpretation of the ellipsis site suggests that NumP is not included within the ellipsis site (see Yoshida et al., 2012 for a related observation).

(48) John read Bill's three books, and Mary read [$_{DP}$ Susan's [$_{NPE}$ ~~(plural) books~~]].

It has been suggested that when a functional head is included in an ellipsis site, the functional

---

[24] It is possible that NumP cannot be elided, for some independent reason, but we are not aware of any compelling reason why we would not assume that NumP, like other phrases, can be elided.

head within the ellipsis site and within the antecedent of ellipsis is subject to a parallelism requirement, or *ellipsis identity* effect (Merchant, 2013b). For example, Merchant (2008, 2013b) suggested that when the Voice head (Voi) is included within the ellipsis site in the VP ellipsis context, parallelism in terms of voice must be observed. Thus, pseudogapping, which Merchant argues to be VoiP-ellipsis, does not tolerate the voice mismatch as in (49b), but VP ellipsis, where VoiP is not included within the ellipsis site as in (49d) does (for other cases of *ellipsis identity* effects, see Merchant, 2008).

(49) a. Some brought roses, and others did [~~$_{VoiP}$ [$_{Voi'}$ Voi [$_{VP}$ buy]]~~] lilies.

b. *Some bought roses, but lilies were [~~$_{VoiP}$ [$_{Voi'}$ Voi [$_{VP}$ [$_{VP}$ bought]]~~] by others.

c. Bill shouldn't remove the trash—the janitor should [$_{VoiP}$[$_{Voi'}$ [Voi [~~$_{VP}$ remove the trash]~~]]].

d. The janitor must remove the trash whenever it is apparent that it should [$_{VoiP}$ [$_{Voi'}$ [Voi ~~[$_{VP}$ be [$_{VP}$ removed]]]]].~~

Similarly, we expect that number in (48) is subject to the *ellipsis identity* effect if Num-head is included within the ellipsis site. However, in (48), parallelism in terms of number is not enforced and a mismatch in number is always tolerated, i.e., the number of the books that John read is not necessarily the same as the number of books that Mary read in (48). This suggests that NumP is not included within the ellipsis site. If NPE is licensed by the Num-head, this pattern naturally follows: The Num-head is located outside the NPE-site, and adjacent to the NPE-site.

3.3.2. Ban on Wh-movement out of the NPE-site

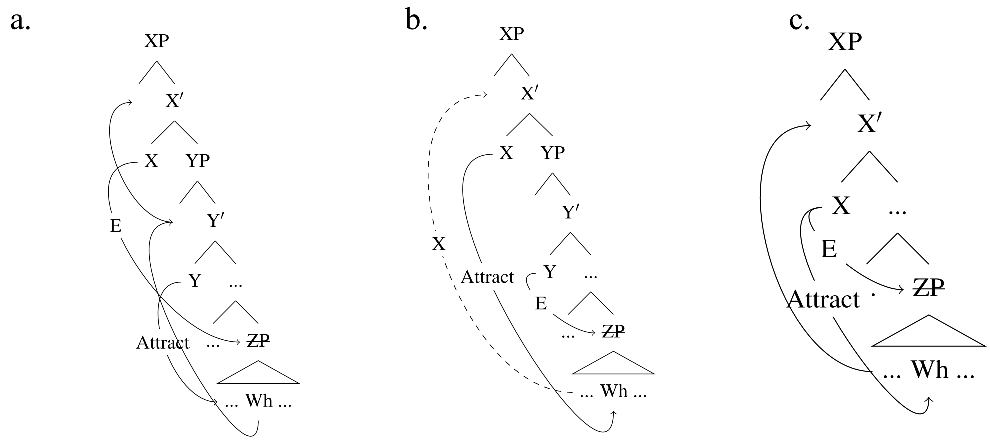
First, we assume derivational ellipsis-licensing i.e., upon the merger of the ellipsis licensing head (the E-head), the ellipsis is licensed. Following Aelbrecht (2010:105), we assume that when a phrase is marked for ellipsis, it is marked for non-pronunciation at PF (or spelled-out to PF), and as a result, syntactic operations cannot affect the constituent marked for ellipsis. For this licensing, we adopt Merchant (2001)'s [E]-feature theory; namely that the ellipsis-site is licensed when the E-feature on the ellipsis licensing head is checked. Second, we assume the theory of phases that the phase head attracts a moving element to the phase-edge (Chomsky, 2008). Furthermore, we assume that the phase head or the head that attract wh-phrase and E-head can be different (Aelbrecht, 2010, 2016).[25] Thus, it is possible that the phase head or wh-attracting head is the E-head, but it is also possible that there is an independent functional head, which is not the phase head/wh-attracting head, that licenses ellipsis. Finally, we adopt the parallel derivation and derivational simultaneity, i.e., when a functional category comes with multiple features in the derivation, the probing of all the probe features start simultaneously, and all the computations run in parallel simultaneously

---

[25] Aelbrecht (2016: 472) says "... in some ellipses it is a non-phase head that establishes the Agree relationship and marks the phase for non-pronunciation."

(Hiraiwa, 2005:45). With these assumptions, we argue that there can be at least three different derivations of ellipsis and extraction out of the ellipsis site illustrated in (50).

(50)



In (50), ZP is the ellipsis site, and E stands for the E-feature checking. The first possibility (50a) is that the E-head X is generated higher than the wh-attracting head (phase head) Y. In this case, derivationally, when Y is merged with the ellipsis site ZP, Y agrees with the wh-phrase and attract the wh-phrase to Spec_YP. The E-head X merges with YP and licenses the ellipsis site ZP at a distance via Agree relation (Aelbrecht, 2010)[26]. Thus, when the ellipsis-site ZP is licensed, the wh-phrase has already been extracted from ZP. The second possibility (50b) is that, the phase-head X is generated higher than the E-head Y. In this case, once Y is introduced to the derivation, Y licenses the ellipsis site ZP. As ZP is marked for ellipsis and

---

[26] For the derivation of sluicing, Aelbrecht (2010:163) argues that there is a wh-attracting head, FocP, which is generated lower than the E-head, C, and thus the wh-phrase can escape the ellipsis site before TP is marked for ellipsis.

frozen for further syntactic operation, when the phase-head X is introduced into the derivation, wh-phrase cannot move out of ZP. The last possibility (50c) is that the phase head is the E-head. Assuming derivational simultaneity, X agrees with ZP and the wh-phrase, and the licensing of the ellipsis site and the attraction of the wh-phrase to the XP-edge run simultaneously in parallel. The ban on wh-movement out of the NPE-site follows from the following assumptions on DP and NPE: (i) D is the phase head in DP and attracts Wh-phrase to the DP-edge (Valois, 1991), (ii) NPE is licensed by Num-head which is generated lower than D, as we have seen earlier.

Let us consider the derivation of an NPE. First, Num, the NPE-licensor bearing the [E]-feature, is merged with NP. Upon merging, the [E]-feature can be checked, and NP is marked for ellipsis and frozen.

(51) [$_{\text{NumP}}$ those [$_{\text{Num}}$ ø$_{\text{_plural}}$] [$_{\text{NP+E}}$ story about who]]

This is why wh-movement out of the NPE-site is not allowed. Basically, when D, the phase-head which attracts the wh-phrase, is introduced in the derivation, wh-phrase is trapped in the ellipsis-marked NP as in (52a). If the ellipsis-licensing does not *take place* derivationally (Aelbrecht, 2010:155), or ellipsis is licensed by the phase-head, D, then wh-movement from the NPE-site should be possible. This is because when D is introduced in the derivation, the NP that contains the wh-phrase *who* is not marked for ellipsis yet, and the D-head can attract *who* to the DP-edge position, as illustrated in (52b).

(52) a. Step 1. [$_{\text{NP}}$ N stories [$_{\text{PP}}$ about who]]

Step 2. [$_{\text{NumP}}$ those [$_{\text{Num'}}$ Num$_{\text{+plural}}$ [$_{\text{NP+E}}$ [N stories [$_{\text{PP}}$ about who]]]]]

Step 3. [$_{DP}$ D [$_{NumP}$ those [$_{Num'}$ Num$_{+plural}$ [$_{NP+E}$ [N stories [$_{PP}$ about who]]]]]]

x

b. Step1. [$_{NP}$ N stories [$_{PP}$ about who]]]

Step 2. [$_{NumP}$ those [$_{Num'}$ Num$_{+plural}$ [$_{NP+E}$ [N stories [$_{PP}$ about who]]]]

Step 3a. [$_{DP}$ those D [$_{NumP+E}$ t Num[$_{NP}$ N stories [$_{PP}$ about who]]]]]

Step 3b. [$_{DP}$ who [$_{DP}$ those D [$_{NumP+E}$ t Num[$_{NP}$ N stories [$_{PP}$ about t]]]]]

The present analysis further predicts that if movement of a phrase is not attracted by D or a phrase can move to a lower position than the DP, movement out of the NPE-site is possible. Yoshida et al., (2012) argued that an NG remnant escapes the NPE-site via *rightward* movement, and to a position lower than NumP[27]. Specifically, they argue that it is adjoined to NP. They show two pieces of evidence for the movement of the NG remnant. First, the NG remnant can co-occur with numerals as in (53). Thus, the ellipsis site in such an example must be lower than NumP. Second, in NG, P-stranding is not possible unlike leftward movement, which is argued to be a property of rightward movement (Jayaseelan, 1990; Lasnik, 1999; Ross, 1967).

(53) *John's three books of music from Penguin Books and [$_{DP}$ Mary's two ~~[$_{NP}$ [$_{PP}$ of t$_{NP}$]~~

[27] We are agnostic about how the rightward movement is to be analyzed ultimately (see Hankamer & Sag, 1976; Kayne, 1994; Larson, 1988; Sag & Hankamer, 1984 for discussion on rightward movement). The important point for us here is that the NG remnant is located lower than the DP or NumP.

~~from Penguin Books]~~ Philosophy] are both highly recommended.

The present analysis predicts that such extraction is possible. Even though what triggers the *rightward* movement is not clear, in (53), the PP should move and adjoin to the NP before or at the point when the Num-head is introduced. Thus, this movement can take place before or at the same time when the NP is marked for ellipsis, resulting in the PP-remnant outside the NPE-site in (53)[28]. Furthermore, in this analysis, the PP remnant of NG escapes from the ellipsis site and thus the PP remnant is not marked for ellipsis. Therefore, we predict that overt extraction from the NG remnant should be possible. The following contrast suggests that it is indeed possible.

(54) a. Mary knows who he wrote these slanderous articles about, but not who he wrote [$_{DP}$ those D [$_{NumP}$ ~~[$_{NP}$ N slanderous stories~~ [$_{PP}$ about t$_{_who}$]]]].

b. *Mary knows who he wrote these slanderous articles about, but not who he wrote [$_{DP}$ those D [$_{NumP}$ ~~[$_{NP}$ N slanderous stories [$_{PP}$ about t$_{_who}$~~]]]].

---

[28] One complication in this analysis is that the lower NP-segment is marked for ellipsis but not the higher one in a configuration like [$_{NumP}$ Num [$_{NP}$ ~~[$_{NP}$ N t$_{_PP}$ PP]~~ PP]]. We would like to tentatively suggest that when segments of a category (Chomsky, 1986) are made by adjunction, any of the segment can be the target of ellipsis (see Baltin, 2012; Hornstein,1994, Merchant, 2000, 2001; Yoshida, 2010 for similar analyses assuming the deletion of a segment of a category).

Following Aelbrecht's (2010) suggestion, we assume that ellipsis is understood as non-pronunciation at PF, which is equivalent to PF-deletion theories (Aelbrecht, 2010; Baltin, 2012). This approach predicts that elements can move *covertly* out of the ellipsis site, in a manner that does not affect the PF-representation. It seems that this is indeed the case. As we have seen when the NPE-site contains a quantifier, this quantifier can scope out elements outside the ellipsis site (and as in the experiment; Table 30). If the inverse scope reading is a result of covert Quantifier Raising (QR: Bruening, 2001; Collins, 2018; Fox, 2000; May, 1977, 1985) it does not affect the order of the constituents in PF. This suggests that QR is indeed possible out of the ellipsis site.

3.3.3. Greek NP-Ellipsis and Extraction

The analysis that we have constructed so far suggests that wh-movement out of the NPE-site is not possible because the NPE-site is licensed by a functional head generated lower than the phase head D, which attracts the overtly moving phrase. This predicts that, in the context where D licenses the ellipsis of the nominal projection, wh-movement or A-bar extraction is possible, i.e., the derivation illustrated in (50c). We attempt to show that the correlation between the ellipsis licensing head and extractability is attested in NPE in Greek. Merchant (2014:20) shows that extraction from the NPE-site in Greek is possible. He cites the following type of examples[29].

[29] Because *the chair* is an episcene noun, with the same masculine and feminine form, we replaced it with *the teacher*.

(55) Tis istorias idhatonpalio [daskalo __], kai [tis glossologias]$_1$

The historian-GEN I.saw the.M old.M teacher.M,and the linguistics-GEN

tha dho ton kenurio [$_{NP}$ ~~daskalo __$_1$~~].

FUT I.see the.M new.M teacher.M (Merchant, 2014: 20)

Lit: "I saw the former teacher (masc) of the history department, and of linguistics, I'll see the new (further attempt to show that the masc) (one)."

In (55), the DP *tis glossologias* is extracted out of the NPE-site. The gender of the elided noun and the antecedent are matched, as the agreement morphology on the adjective indicates the gender marking of the elided noun. Merchant further shows that when the gender specification of the elided noun and the antecedent are not matched, extraction out of the NPE-site gives rise to the degraded acceptability, as illustrated in (56).

(56) Tis istorias idhatonpalio [daskalo __], kai [tis glossologias]$_1$

The historian-GEN I.saw the.M old.M teacher.M,and the linguistics-GEN

tha dho tin kenuria [$_{NP}$ ~~daskala __$_1$~~].

FUT I.see the.F new.F teacher. F (Merchant, 2014: 21)

Lit: "I saw the former teacher (masc) of the history department, and of linguistics, I'll see the new (female) (one)."

Merchant argues that this extraction asymmetry follows if it is the case that the gender matching (55) involves ellipsis but gender mismatching (56) involves an empty pro-form. Although Merchant's analysis elegantly captures the extraction asymmetry and other differences between gender matching and mismatching NPEs, we would like to suggest a potential alternative analysis extending the analysis of NPE in English.

The gender match/mismatch can be understood as an *ellipsis identity* effect (Merchant, 2008, 2013b, 2014). As we have discussed earlier, when a functional head is included inside the ellipsis site, the elided functional head is subject to the *ellipsis identity* requirement. In other words, when the gender is matched, the functional head that is responsible for gender (let us call it X) is included within the ellipsis site. On the other hand, when the gender is mismatched, the functional head responsible for gender is excluded from the ellipsis site. Thus, it is possible that the ellipsis site is larger (including a functional head responsible for gender) when the gender is matched, but it is smaller (excluding a functional head responsible for gender) when the gender is mismatched. In Greek NPE, therefore, when the gender is mismatched, the ellipsis site is licensed by the functional head responsible for gender agreement, which is generated lower than the D head. In this case, gender information is not subject to the *ellipsis identity* requirement and extraction out of the ellipsis site should not be possible, as illustrated in (57). When the phase head D is introduced, the ellipsis site checks the E-feature of the functional head, the NP is marked for ellipsis, and thus syntactic operations cannot affect the constituent marked for ellipsis (Aelbrecht, 2010:105). Anything that is contained in this NP cannot move out overtly.

(57) Step 1. [$_{NP}$ [$_{N}$ daskala [$_{DP}$ tis glossologies]]]

Step 2. [$_{XP}$ X [$_{NP\_+E}$ [$_{N}$ daskala [$_{DP}$ tis glossologies]]]]

On the other hand, when the gender is matched, the ellipsis site includes the functional head responsible for gender agreement, and thus it is subject to the *ellipsis identity* requirement.

In this case, the ellipsis site should include the gender-related functional head, and thus, the ellipsis site should be licensed by some other head. We contend that it is D, the phase head. If so, the extraction should be possible because D can attract the moving constituent and license the ellipsis site at the same time as illustrated in (58).

(58) Step 1. [$_{NP}$ [N' daskalo [$_{DP}$ tis glossologies]]]

Step 2. [$_{XP}$ X [$_{NP\_+E}$ [N' daskalo [$_{DP}$ tis glossologies]]]]

Step 3a. [$_{DP}$ D [$_{XP\_+E}$ X [$_{NP}$ [N' daskalo [$_{DP}$ tis glossologies]]]]]

Step 3b. [$_{DP}$ [$_{DP}$ tis glossologiess] [$_{D'}$ D [$_{XP\_+E}$ # [$_{NP}$ [N' daskalo [t_$_{DP}$]]]]]

In this way, the proposed analysis correctly predicts the patterns of extraction from the NPE-site in Greek. However, is there any piece of evidence that suggests gender mismatching NPE in Greek involves ellipsis? As we have discussed earlier, in the ellipsis construction, the quantifier embedded within the ellipsis site can out-scope an element outside the ellipsis site (Collins, 2015, 2018; Merchant, 2013a:539). The inverse scope interpretation (*for every type of music, there are some singers who are bad teachers for that type of music*) is available in NPE in gender matching NPE as in (59)[30].

(59) Kapji kitharistes ine kali daskali

Some.M.NOM.PL guitarists.M.NOM.PL. are-3PL good.M.NOM.PL teachers.M.NOM.PL

se kathe idos mousikis ala kapji

[30] We thank Emilia Molimpakis for help with the Greek examples and judgments.

in every.N.ACC type.N.ACC music.F.GEN but some.M.NOM.PL

trajoudistes ine kaki

singers.M.NOM.PL are-3PL bad.M.NOM.PL

~~[NP [daskali [PP [P se [DP kathe idos mousikis]]]]]~~

teachers.M.NOM.PL in every.N.ACC type.N.ACC music.F.GEN

"Some guitarists are good teachers of every type of music, but some (male) singers are bad."

The inverse scope interpretation (*for every type of legal subject, there are some friends of mine who are bad lawyers for that type of legal subject*) is also available in gender mismatching NPE, where *dikijoros* (the lawyer) concides with both genders;

(60) Kapjios filos mou ine kalos dikijoros se

Some.M.NOM.PL friend.M.NOM mine-GEN is-3SG good.M.NOM lawyer.NOM in

kathe nomiko andikimeno ke kapoios fili

every.N.ACC legal.N.ACC subject.N.ACC and some.F.NOM friend.F.NOM

mou ine kaki ~~[NP [dikijoros[PP [P se [DP kathe nomiko andikimeno]]]]]]~~

mine.GEN is-3SG bad.F.NOM lawyer.NOM in every.N.ACC legal.N.ACC subject.N.ACC

"Some friend of mine is a good lawyer in every law subject and some female friend of mine is bad."

The inverse scope interpretation is available for the non-ellipsis counterparts of these examples as well. All these examples thus suggest that both the gender matching NPE and the gender mismatching NPE involve ellipsis.

3.3.4. English VP-Ellipsis and Wh-extraction

If the size of the ellipsis sites correlates with extractability as suggested in Greek NPE, we expect that if there are two possible ellipsis sites within the same phase domain, and if the lower phrase is elided, overt A-bar extraction from the ellipsis site is banned; but, if the higher phrase is elided, overt A-bar extraction from the ellipsis site should be possible. It seems that VPE and VoiceP-ellipsis (VoiPE: Merchant, 2013b) in English show exactly these patterns i.e., English (Voice-mismatching) VPE disallows wh-extraction out of ellipsis site, unlike English VoiPE. It has been known that English VPE allows A-bar extraction from the ellipsis site (Aelbrecht, 2010; Hartman, 2011; Merchant, 2013; Schuyler, 2001; Takahashi & Fox, 2005)[31].

(61) I remember which box Steve didn't send to Mary, but I know which box Steve did

---

[31] The logic of MaxElide is, if there is a larger constituent that can be elided, then it should be elided, if the ellipsis site contains A-bar trace (Lasnik & Park, 2013). In our case, TP is the largest elidable material.

(1) Mary said he wrote these stories about someone, but she cannot tell who.

The NP is properly contained within TP, and thus if any element is to be elided, TP should also be elided. However, when there is an element in the ellipsis hosting clause that is contrasted to an element in the antecedent clause, the contrasted element cannot be elided, therefore the largest deletable constituent is now NP like in (2).

(2) Mary can tell who he wrote THESE stories about $t_{\_who}$, but she cannot tell who he wrote THOSE [$_{NP}$ ~~stories about t~~ $_{\_who}$].

~~[send t to Mary].~~

When the ellipsis site and the antecedent mismatch in voice, wh-extraction is not possible:

(62) *I remember which box Steve didn't send to Mary, but I know which box was ~~[sent t to Mary]~~.

Note that the active voice can antecede the passive ellipsis in the VPE context (i.e., Merchant 2008, 2013b) as in (63a), and wh-movement is possible in the non-ellipsis counterpart of (62), as in (63b).

(63) a. I remember that Steve didn’t send this box to Mary, but I know that this box was ~~{sent to Mary}.~~

b. I remember which box Steve didn't send $t_{\text{which box}}$to Mary, and I know which box was sent to $t_{\text{which box}}$ Mary.

The contrast between (61) and (62) highlights the correlation that we have pointed out above. First, the voice match/mismatch effect can be understood as the *ellipsis identity* effect (Merchant, 2013b). As Merchant (2013b:89) has argued, when voice is mismatched, the voice head is not included within the ellipsis site. Thus, a constituent smaller than VoiP, namely VP, is elided. When, on the other hand, voice is matched, the voice head is included within the ellipsis site. In this case the VoiPE should be licensed by a functional head that is generated higher than VoiP, which we take as v (Merchant, 2013b), the phase head (the possibility illustrated in (50c)). Thus in (63), VoiP is elided when v licenses the VoiPE-site. When the ellipsis site includes VoiP which is responsible for voice, voice is subject to the *ellipsis identity* effect, and the voice of the antecedent and the ellipsis site must be matched. In this case, the ellipsis site should be bigger, including the voice-related functional head, and thus, the ellipsis must be licensed by v, the phase head. If so, the extraction should be possible because when v is introduced in the derivation, the VP that contains the wh-phrase *which* is not marked for ellipsis, and v can agree with *which box* and attract it to vP-edge.

(64) a. Step 1. [$_{\text{VP}}$ V sent [$_{\text{PP}}$ to Mary which box]]

Step 2. [$_{\text{VoiP}}$ Voice [$_{\text{VP+E}}$ [V sent [$_{\text{PP}}$ to Mary which box]]]]]

Step 3. [$_{\text{vP}}$ v [Voice [$_{\text{VP+E}}$ [V sent [$_{\text{PP}}$ to Mary which box]]]]]

x

b. Step 1. [$_{\text{VP}}$ V sent [$_{\text{PP}}$ to Mary which box]]

Step 2. [$_{\text{VoiP}}$ Voice [$_{\text{VP+E}}$ [V sent [$_{\text{PP}}$ to Mary which box]]]]]

Step 3a. [$_{\text{vP}}$ [$_{\text{v'}}$ v [$_{\text{VoiP}}$ Voice [$_{\text{VP+E}}$ [V sent [$_{\text{PP}}$ to Mary which box]]]]]]

Step 3b. [$_{\text{vP}}$ [which box][$_{\text{v'}}$ v [$_{\text{VoiP}}$ Voice [$_{\text{VP+E}}$ [V sent [$_{\text{PP}}$ to Mary t]]]]]]

However, in (62), the voice is mismatched. This means that the ellipsis site does not include VoiP and thus the ellipsis site is licensed by the functional head responsible for voice, which is generated lower than v. Wh-extraction is impossible here because when the Voi-head is merged to VP, VP is marked for ellipsis (the possibility illustrated in (50b)). Thus when the phase head v is introduced, VP is marked for ellipsis and the wh-phrase embedded within VP cannot overtly move out. Thus, the analysis that we have proposed can be corroborated by the wh-extraction pattern in the VP/VoiP-ellipsis context.

So far, we have assumed that NPE in English is licensed by the Num-head but not by the D-head, but why can D not be a possible ellipsis licensor in English as in Greek? There are two possibilities that we would like to consider. First, unlike other functional elements in nominal phrases in English, articles like the definite determiner *the*, the indefinite determiner *a/an*, and the possessive/genitive marker *'s* do not impose any agreement morphology on the noun or Number-head, or the determiners themselves. If agreement morphology is necessary for a nominal phrase to be elided, (Kester, 1996; Llombart-Huesca, 2002; Lobeck, 1995), it follows that English determiner, which does not impose any agreement morphology, cannot license ellipsis. Another possibility is that D in English behaves more like a clitic that is attached to the NumP or NP. When ellipsis occurs, there is no host for the determiner. On the other hand, the apostrophe, *'s,* is enclitic and it needs a host to the left. Thus, when there is a DP, *'s* can be cliticized to that DP. In this approach, an obvious problem is that *the/a* cannot co-occur with Nominal Gapping, where the PP remnant is right-adjacent to the D. One possible analysis is that *the* and *a* are clitics that need to be attached to NP/NumP. In cases where these elements are either displaced, deleted or empty, it cannot cliticize to it. This position is supported by the following observations. English Ds cannot appear in a right-node raising context as in (65a), suggesting that they must have NumP/NP adjacent to them.

(65) a. *The and a book that Mary bought.

b. These and those books that Mary bought.

Ds in English are not morphologically independent and require the overt host adjacent.

3.4. Summary of the Syntax of NPE

We argued that the NPE-site is associated with silent syntactic structure that holds a certain parallelism with the antecedent NP by examining the BCC connectivity effect and inverse scope effects in the NPE-context. We also argued that wh-movement out of an NPE-site is banned not because the NPE-site is associated with an empty pro-form, but because the head that licenses the NPE-site and the head that attracts the wh-phrase (the phase head within DP-domain) are different, and because the NPE-licensing head is generated lower than D, the phase head. Thus, the NP that contains the launching site of wh-movement undergoes ellipsis before wh-movement within the DP takes place. This predicts that, in the context where D licenses the ellipsis of the nominal projection, wh-movement or A-bar extraction is possible. We further attempt to show that the correlation between the ellipsis licensing head and extractability is attested in NPE in Greek, and in VP-Ellipsis/VoiceP-Ellipsis in English. This suggests that the NPE-site involves hierarchical syntactic structures that parallel those of the antecedent NP in contrast to Anaphoric *one* and Pronoun *it*. Wh-movement out of an NPE-site is prohibited not because the NPE-site involves an empty pro-form. Rather, the NPE-site is marked for ellipsis and frozen for syntactic operations by the ellipsis licensing head prior to the introduction of the phase head into the derivation, which attracts the wh-phrase to the phase edge. We have shown that the NPE-site is associated with a rich hierarchical structure. The recovery of the content of ellipsis indicates the recovery of the structure within the NPE-site. The recovery of the content of the ellipsis site crucially involves the reference to the antecedent of the ellipsis site. A detailed study of the syntax of NPE will serve as a basis for us to uncover the processes of identifying and recovering the content of the NPE-site in regard to antecedent retrieval.

3.5. Introduction: Antecedent Retrieval

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 (66).

(66) 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 recognize 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 (67), *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 (67) is processed, the parser needs to *recover* content into the NPE-site by referring to the content of the antecedent.

(67) 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 (68) (Hornstein & Lightfoot, 1981, Lidz, Waxman, & Freedman, 2003; Pearl & Lidz, 2009, among others). Thus, it is plausible that when an Anaphoric *one* is processed, the parser accesses and retrieves the antecedent of the Anaphoric *one*.

(68) 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.

3.5.1. Memory Retrieval Mechanism

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 et al., 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 et al., 2013; Lago et al., 2015; Parker & Phillips, 2017; Pearlmutter, et al., 1999; Tanner et al., 2014; Thornton & MacDonald, 2003; Wagers et al., 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, et al., 1997; Parker & Phillips, 2017; Pearlmutter et al., 1999; Tanner et al., 2017; Tanner et al., 2014; Thornton & MacDonald, 2003; Wagers et al., 2009). For example as in (69), 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.

(69) *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 et al., 2011; Sturt, 2003; Tanner et al., 2014).

Another point in favor 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 characterized 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 & Dosher, 1989; McElree et al., 2003; 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 et al., 2008; Pearlmutter et al., 1999; Parker & Phillips (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 (70c) and (70d) are ungrammatical, (70d) is read faster at the verb region and rated more acceptable than (70c) 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 (70a) and (70b) (Wagers et al., 2009).

(70) 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.

3.5.2. Agreement Attraction as a Probe

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.

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 recognized.[32]

(71) 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.

---

[32] 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, Clifton, Frazier, & Moulton, 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.

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 (71), once the NPE-site is processed, the parser first retrieves the head noun (*[$_{head\text{-}N}$ 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 (71a-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, (71b), then the processing of the verb would be facilitated. On the other hand, if they do not match in number, (71a), 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, (71c) and (71d) as the agreement is successfully licensed at first pass.

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; Gordon, Hendrick, Johnson, & Lee, 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 (71d) and ungrammatical (71a) NPE cases.

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 71b/71d vs 71a/71c. However, reading times will be slower in both ungrammatical conditions (71a and 71b) 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 judgments 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.

3.5.3. Motivation for the Experiments

In order to disclose what is retrieved during the processing of elided and anaphoric elements, 12 experiments were designed to contrast agreement attraction in NPE with overt

NPs (Experiments 6) and nominal anaphora (Anaphoric *one* and Pronoun *it;* Experiment 8 and 9). These included 4 offline acceptability rating experiments (Experiments 6a, 7a, 8a, and 9a), 4 self-paced moving window reading experiments (Experiments 6b, 7b, 8b, and 9b) and 4 eye-tracking while reading experiments (Experiments 6c, 7c, 8c, and 9c).

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 6a and faster reading times for ungrammatical sentences with plural local nouns (vs singular local nouns) in Experiment 6b. 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 6a) and read faster (Experiment 6b & 6c) 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 6a) or in reading times (Experiment 6b & 6c).

The goal of Experiment 7 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 (72).

(72) 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 6, 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 7a and to similar reading times in ungrammatical sentences with plural versus singular local nouns in the No Anaphora condition in Experiment 7b & 7c. 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 7a and read faster in Experiment 7b & 7c than their singular counterparts (*key to the cell… necklace are*).

The goal of Experiment 8 was to rule out a final alternate account of the data, testing whether the agreement attraction observed in NPE contexts in Experiment 6 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 8 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 6. 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, 1991; Hornstein & Lightfoot, 1981; Lidz, Waxman, & Freedman, 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, 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 6. 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.

The goal of Experiment 9 was to further examine the nature of the antecedent retrieval mechanism. The content of the Pronoun *it* is also supplied by its antecedent, meaning that its interpretation is dependent on its antecedent. However, in contrast to NPE, Pronoun *it* is known to have no internal structure (Hankamer & Sag, 1976; c.f., Elbourne, 2005; Postal, 1969). This raises an interesting question: in NPE cases, the parser first builds the structure and inspects an agreement. If it fails, the parser accesses the features in a parallel manner. In this sense, for Pronoun *it*, morphological cues may have a priority over structural information as a retrieval cue to retrieve the information when the pronoun is encountered. Pronoun *it* does not probe the plurality as it signals anaphoric NP and a singular NP. Thus, Pronoun *it* can be a good testing ground to reveal how and when morphological and syntactic information are at play in resolving such dependencies. We predict the reading times facilitation and increased acceptability ratings of the plural local noun in comparison to the singular local noun in ungrammatical conditions for the baseline conditions, but not for Pronoun *it*. We predict grammaticality effects during early measures at the earlier stage of processing, but no agreement attraction effect in later measures in an eye-tracking while reading experiment such as RPD or TFT. Therefore, the lack of agreement attraction for Pronoun *it* would suggest that although the presence of *it* predicts the upcoming verb, the processing difficulty caused by the mismatch in number can be modulated by the morphological cues to access the antecedent in content-addressable memory. On the other hand, if Pronoun *it* shows agreement attraction, this suggests that the parser is retrieving any antecedent without recourse to the structure. Therefore, testing Pronoun *it* sets up a good testing ground whether the agreement attraction for NPE is truly due to the retrieval of the structural information.

3.6. Experiment 6a NPE: An Acceptability Rating Experiment (Offline)

3.6.1. 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 2×2×2 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 summarized in Table 31. The first five words of each experimental item in the first conjunct always followed the form shown in Table 31 (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 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 minimize 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.

Table 31. Sample stimuli for experiment 6.

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

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.

3.6.2. 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.

3.6.3. Analysis

Data were analyzed with linear mixed effect regression using the *lme4* package in R version 3.2.3 (Baayen, 2008; Baayen et al., 2008; Bates et al., 2014; Jaeger, 2008).[33] Each model

---

[33] As an anonymous reviewer in *Language, Cognition, and Neuroscience* 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 6a. 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* ($\beta = 0.65, SE= 0.19, z= 3.49, p<0.001$).

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 et al., 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).

3.6.4. Results

Mean acceptability scores are shown in Table 32, and mixed effect model outputs are shown in Table 33. 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 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* (*β = 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.

Table 32. Mean acceptability ratings from experiment 6a.
(Standard errors are in parentheses)

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)

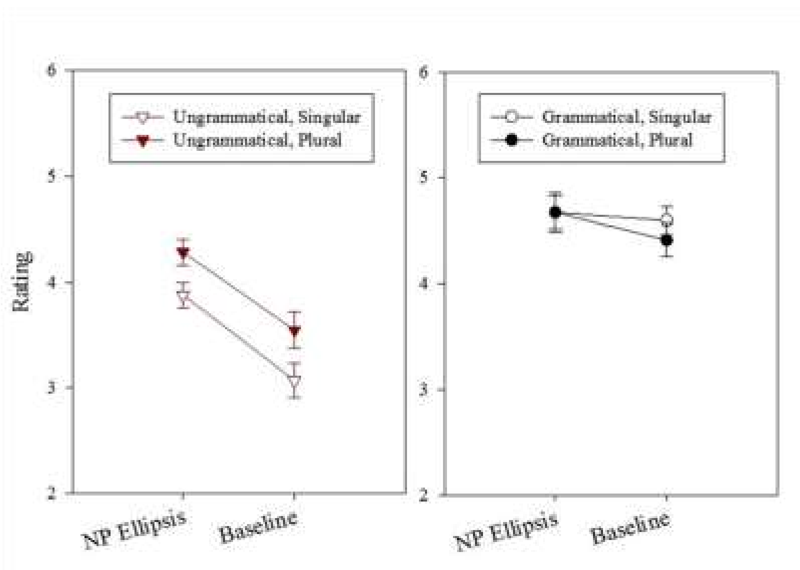


Figure 27. Mean acceptability ratings from experiment 6a.
Error bars indicate standard error.

Table 33. Summary of fixed effects from linear mixed effects model in experiment 6a. Random intercepts were included for subjects and items, as were by-subject intercepts for *Local noun*, *Grammaticality*, *NPE* and *Local noun* x *Grammaticality*, and by-item intercepts for *Local noun*, *Grammaticality*, *NPE* and *NPE* x *Grammaticality*.

	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 x *Grammaticality*	0.52	0.18	2.95	$< 0.01^{**}$
Grammaticality x *NPE*	0.65	0.19	3.42	$< 0.01^{**}$
Local noun x *NPE*	0.06	0.15	0.39	0.70
Local noun x *Grammaticality* x *NPE*	-0.27	0.29	-0.91	0.37

3.7. Experiment 6b NPE: A Self-Paced Word-by-Word Moving Window Experiment

3.7.1. 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 6a were used (see Table 31); 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.

3.7.2. 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, et al., 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 familiarized with the procedure. The experiment took each participant an average of approximate 30 minutes to complete.

3.7.3. Analysis

Following Kazanina et al., (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 6a. The regions used for analysis consisted of single words. The critical regions were the verb, the following word (spillover region 1) and one word after the spill over region 1 (spillover region 2).

3.7.4. Results

The region-by-region reading times for baseline conditions are presented in Figure 28; those for NPE sentences are presented in Figure 29. Reading times at the critical spillover region for both NPE and baseline conditions are presented in Figure 30 and mixed effect model outputs are presented in Table 34. Mean accuracy for critical trial comprehension questions was 80.0%.

Table 34. Summary of results of linear mixed effects models by region in experiment 6b.

	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 x *Grammaticality*	-5.18	7.73	-0.67	0.50
Grammaticality x *NPE*	0.17	7.73	0.02	0.98
Local noun x *NPE*	-7.02	7.73	-0.91	0.36
Local noun x *Grammaticality* x *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 x *Grammaticality*	-25.88	7.33	-3.53	<0.001***
Grammaticality x *NPE*	-22.37	7.32	-3.05	< 0.01**
Local noun x *NPE*	1.78	7.32	0.24	0.81
Local noun x *Grammaticality* x *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 x *Grammaticality*	-2.45	7.13	-0.34	0.73
Grammaticality x *NPE*	-9.18	7.13	-1.29	0.20
Local noun x *NPE*	2.65	7.13	0.37	0.71
Local noun x *Grammaticality* x *NPE*	-15.24	14.26	-1.07	0.29

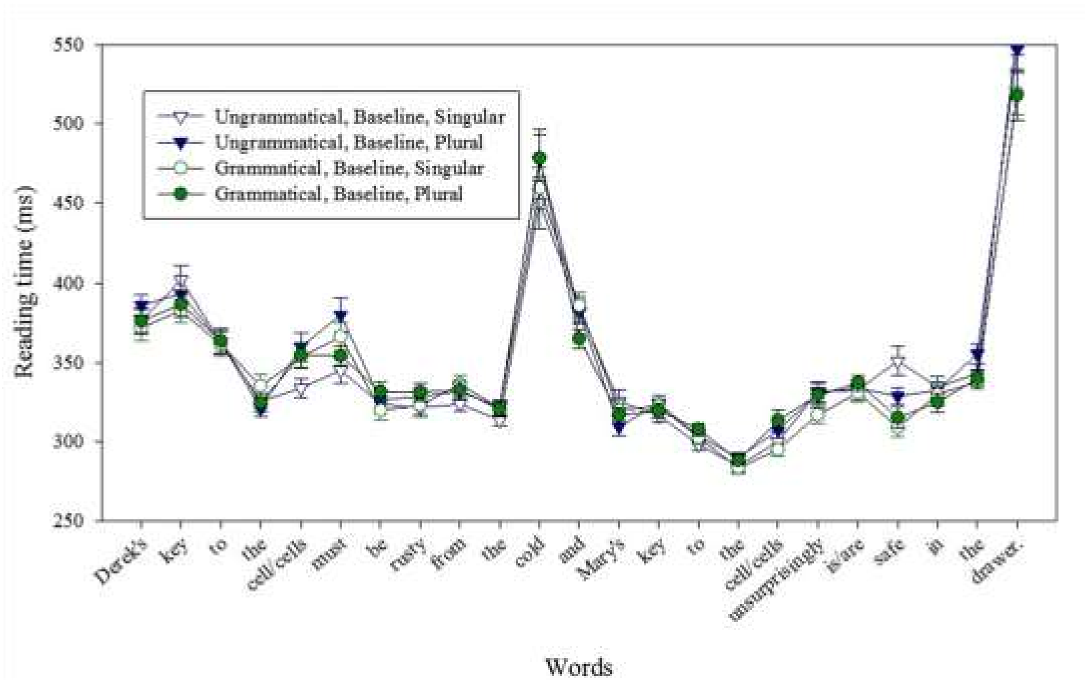


Figure 28. Region-by-region reading time means from experiment 6b for baseline conditions. **Error bars indicate the standard error. The regions of interest are *is/are* (verb), *safe* (spillover 1), and *in* (spillover 2).**

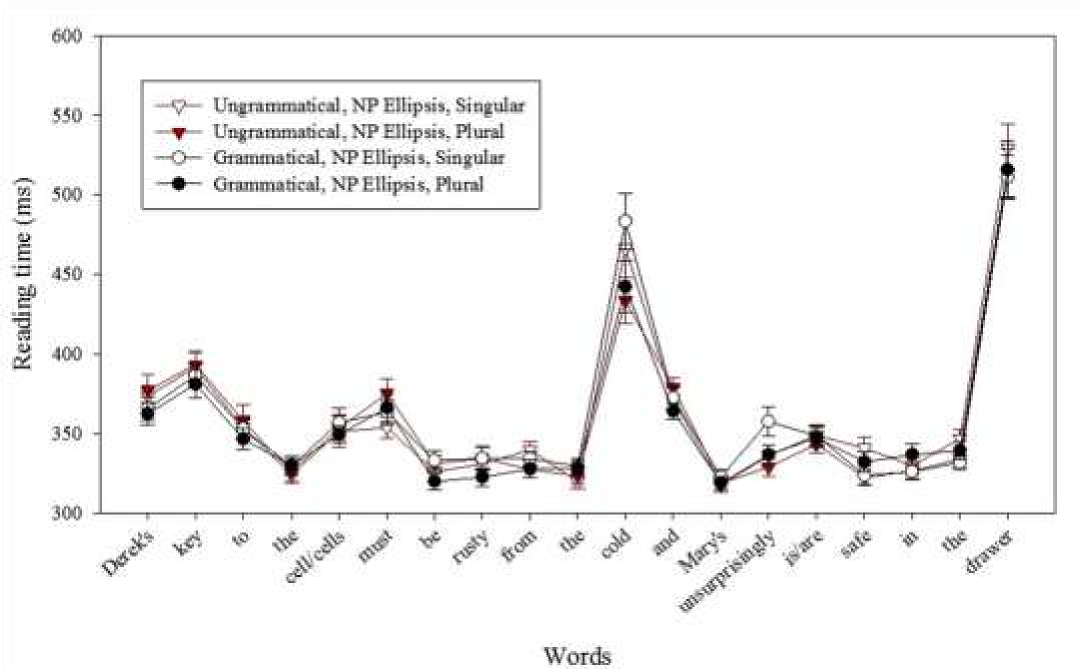


Figure 29. Region-by-region reading time means in experiment 6b for NPE conditions. Error bars indicate the standard error. The regions of interest are *is/are* (verb), and *safe* (spillover 1), and *in* (spillover 2).

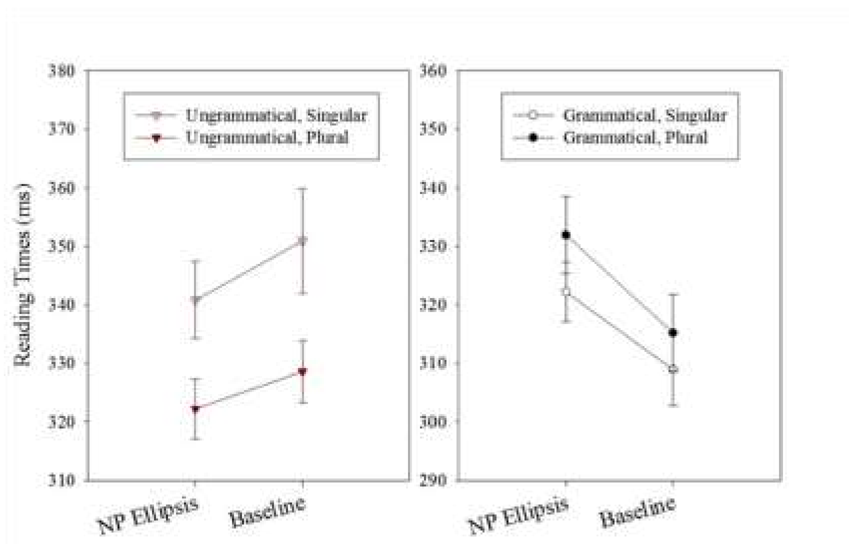


Figure 30. Reading times at the spillover region 1 (*safe*) for all conditions in experiment 6b. **Error bars indicate the standard error.**

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 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.

3.8. Experiment 6c NPE: An Eye-tracking while Reading Experiment

3.8.1. Participants, Materials and Design

In this experiment, 40 native speakers of English from the Northwestern University community with no history of language disorders participated. Before the experiment, participants provided informed consent and got course credit upon participation.

Same critical items were used as in Experiment 6b. Items were distributed in a pseudo-randomized manner so that the experimental items of from the same experiment did not appear adjacent to each other. The experimental items were mixed with 64 filler sentences of similar length and complexity.

3.8.2. Procedure

Participants read each single sentence while their eye movements were tracked using a tower-mounted SR Research EyeLink 1000 eye-tracker, sampling eye-movements at the rate of 2000Hz. Before the experiment began, participants were calibrated with a nine point calibration, and recalibration was conducted whenever necessary. The participants were given short breaks if necessary, but the short break was always followed by the recalibration. At the start of each experimental trial, the left edge of the monitor screen revealed a little black rectangular. When the calibration was successful, the right rectangular disappeared and successful calibration led to the presentation of the experimental item. Participants were asked to answer comprehension questions after each sentence, by pressing right or left button on a game control pad. The overall experiment took around 30-45 minutes.

3.8.3. Analysis

The gathered Eye-fixation data were manually corrected for the purpose of the vertical drift. In this experiment, we focus on the gaze duration at three different regions: the critical region (*is/are*), spillover region 1 (e.g., *rusty from*), and spillover region 2 (e.g., *the cold*). Fixations of less over 1000 ms, and the total fixation time over 4000ms were excluded from the analysis. In this experiment, we focus on three different eye-movement measures which are First Fixation Duration (FFD), Regression Path Duration (RPD), and

Total Fixation Time (TFT). First Fixation Duration (FFD) is the duration of the first fixation in a region which is comprised of the time the region is first entered from the left, until a subsequent fixation is made. Regression Path Duration (RPD) is the sum of fixation durations of the time the region is first entered until the eye moves to the right, often referred to as *go-past time* (Staub & Rayner, 2007). Total Fixation Time (TFT) is the sum of all fixations on the region, including first pass reading and re-reading times. For any given trial, if the measure returned no data (e.g., if there were no fixations on the region), the trial was treated as a missing value in that particular region (Sturt, 2003).

The statistical analysis was carried out with the log-transformed data with the aim of normality (Box & Cox, 1964; Vasishth et al., 2013). Identical dependent measures were used as in 6a. All models involved the maximal random effects structure which fit the data (Barr et al., 2013), including random intercepts for participants and items and random slopes for fixed effects provided the model converged.

3.8.4. Results

<u>Verb Region</u>

In the Regression Path Duration measure, a main effect of *NPE*, an interaction of *Local noun* and *Grammaticality,* as well as a three-way interaction of *Local noun*, *Grammaticality*, and *NPE* were observed. This was further qualified by the subset analysis where a main effect of *Local noun* ($\beta = -0.44, SE= 0.15, t= -2.85, p<0.01$), a main effect of *NPE* ($\beta = 0.67, SE= 0.17, t= 3.93, p<0.01$), and an interaction between *Local noun* and *NPE* were observed in ungrammatical conditions ($\beta = -0.62, SE= 0.29, t= -2.15, p=0.05$). In contrast, only an interaction between *Local noun* and *NPE* was observed in grammatical conditions (*β = 1.13, SE= 0.38, t= 2.94, p<0.01*). Other measures did not reach significance.

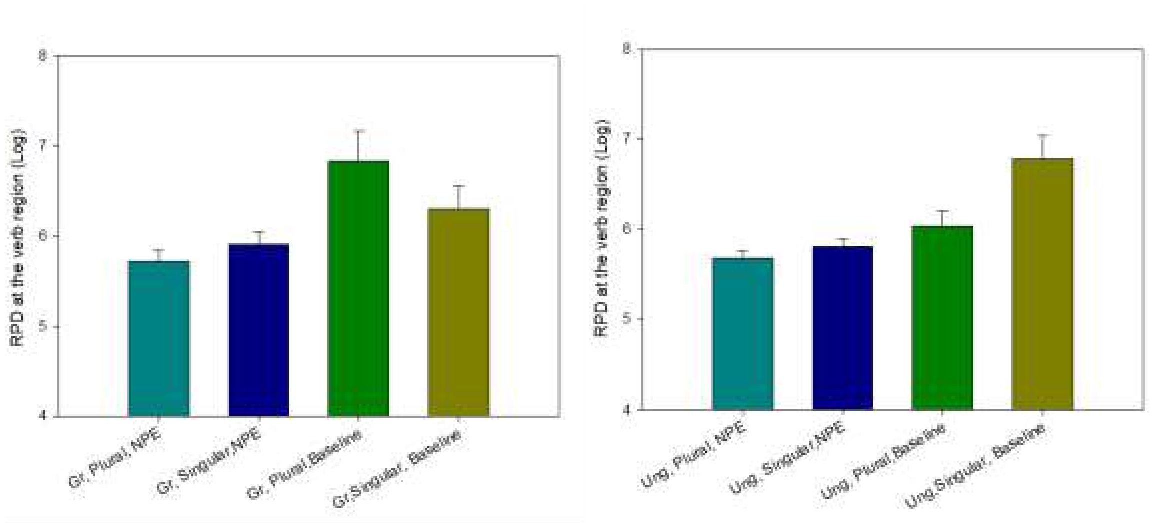


Figure 31. Bar plot for the RPD in grammatical conditions (left) and ungrammatical conditions (right) at the verb region for NPE and the baseline in experiment 6c.

<u>Spillover Region 1</u>

In the First Fixation Duration, a main effect of *NPE* was observed where the baseline constructions were read significantly slower than the NPE constructions. In the Regression Path Duration measure, a main effect of *Local noun* was observed where plural local nouns were read faster than the singular local nouns. We also observed a main effect of *Grammaticality*. This was driven by an interaction between *Local noun* and *Grammaticality* such that ungrammatical plural local nouns were read significantly faster than their counterparts in ungrammatical conditions, but not in grammatical conditions. Further subset analysis revealed a main effect of *Local noun* (*β = -0.45, SE= 0.13, t= -3.48, p<0.01*) and *NPE* (*β = 0.59, SE= 0.16, t= 3.71, p<0.001*) in ungrammatical conditions, but no interaction between *Local noun* and *NPE*. There was only a main effect of *NPE* in grammatical conditions (*β = 0.44, SE= 0.15, t= 2.97, p<0.01*). This suggests that NPE conditions and the baseline conditions were read in a similar manner, suggesting that an illusion of grammaticality is observed. Further subset analysis also revealed an interaction between *Local noun* and *Grammaticality* in the baseline (No NPE) condition (β = *-0.65, SE= 0.25, t=-2.61, p<0.01*), a main effect of *Local noun* (β = *-0.23, SE= 0.10, t= -2.24, p<0.05*), and an interaction between *Local noun* and *Grammaticality* in NPE condition (β = *-0.41, SE= 0.20, t= -2.09, p<0.05*). In the Total Fixation Time duration, we also observed an interaction between *Local noun* and *Grammaticality*.

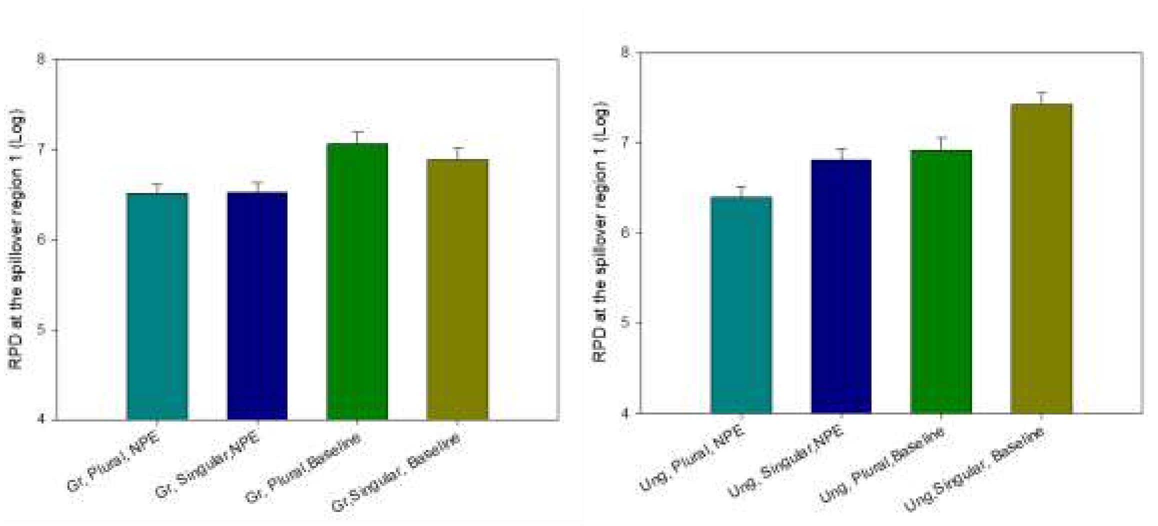


Figure 32. Bar plot for the RPD in grammatical conditions (left) and ungrammatical conditions (right) at the spillover region 1 for NPE and the baseline in experiment 6c.

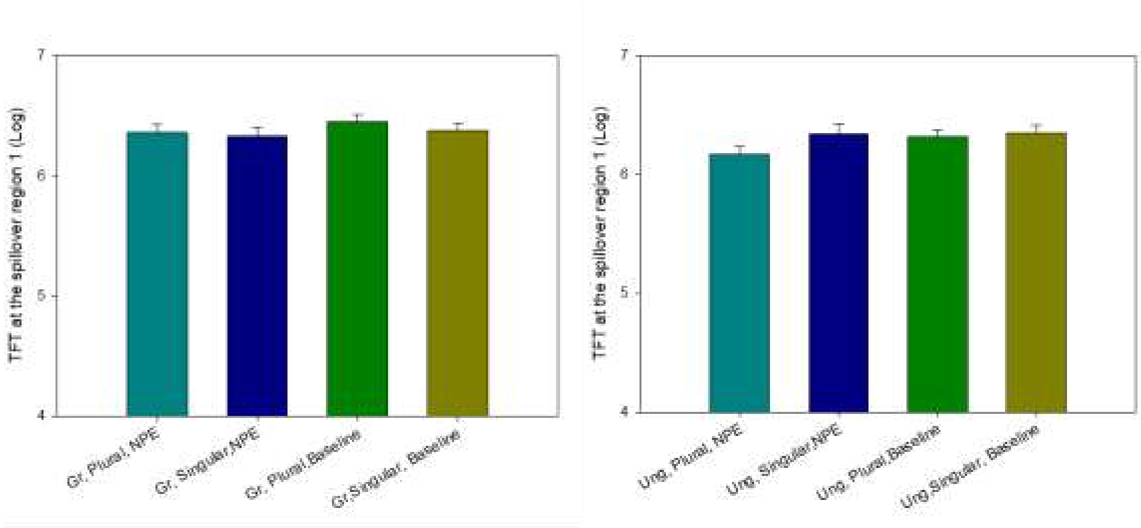


Figure 33. Bar plot for the TFT in grammatical conditions (left) and ungrammatical conditions (right) at the spillover region 1 for NPE and the baseline in experiment 6c.

Spillover Region 2

In the Regression Path Duration, an interaction between *Local noun* and *Grammaticality* was observed. Further subset analysis revealed a main effect of *Local noun* in ungrammatical conditions ($\beta = -0.19, SE= 0.09, t= -2.06, p<0.05$), but not in grammatical conditions. In the Total Fixation Time measure, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower than grammatical sentences. Furthermore, a main effect of *NPE* was observed such that NPE constructions were read significantly slower than the baseline constructions.

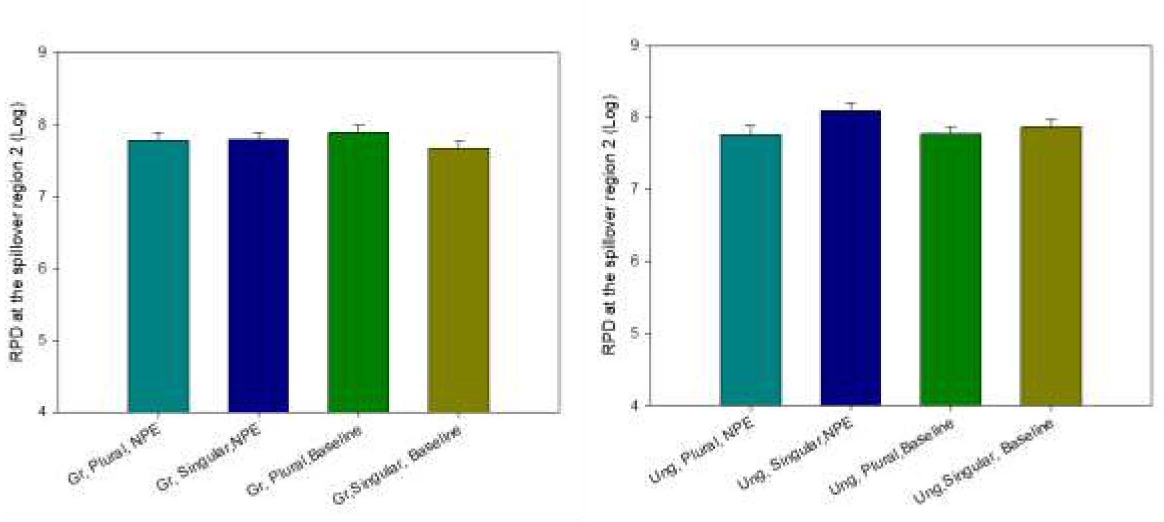


Figure 34. Bar plot for the RPD in grammatical conditions (left) and ungrammatical conditions (right) at the spillover region 2 for NPE and the baseline in experiment 6c.

Table 35. The statistical analysis of results for the eye-tracking while reading experiment (experiment 6c) on the verb region, spillover region 1, and spillover region 2.

	Estimate	SE	t	p
Verb Region (*is/are*)				
First Fixation Duration				
(Intercept)	5.39	0.33	163.83	
Local noun	-0.01	0.05	-0.12	0.90
Grammaticality	0.03	0.04	0.71	0.50
NPE	0.01	0.06	0.11	0.92
Local noun Grammaticality	-0.00	0.08	-0.03	0.98
Grammaticality x *NPE*	0.13	0.08	1.65	0.11
Local noun x *NPE*	-0.06	0.08	-0.73	0.48
Local noun x *Grammaticality* x *NPE*	-0.18	0.16	-1.17	0.25
Regression Path Duration				
(Intercept)	-6.12	0.09	66.55	
Local noun	-0.14	0.12	-1.15	0.26
Grammaticality	-0.12	0.14	0.84	0.42
NPE	0.69	0.16	4.45	<0.001 ***
Local noun Grammaticality	-0.57	0.24	-2.36	<0.05 *
Grammaticality x *NPE*	-0.16	0.24	-0.67	0.51
Local noun x *NPE*	0.11	0.24	0.47	0.64
Local noun x *Grammaticality* x *NPE*	-1.35	0.49	-2.77	<0.01 **
Total Fixation Time				
(Intercept)	5.74	0.05	114.27	
Local noun	0.03	0.09	0.33	0.75

Grammaticality	0.08	0.07	1.09	0.29
NPE	-0.08	0.08	-0.96	0.35
Local noun Grammaticality	0.01	0.12	0.08	0.93
Grammaticality x *NPE*	0.14	0.12	1.14	0.27
Local noun x *NPE*	-0.03	0.12	-0.30	0.77
Local noun x *Grammaticality* x *NPE*	-0.17	0.23	-0.71	0.51
Spill-over region 1 (*rusty from*)				
First Fixation Duration				
(Intercept)	5.36	0.02	251.96	
Local noun	-0.01	0.03	-0.55	0.59
Grammaticality	0.01	0.03	0.41	0.68
NPE	0.07	0.03	2.52	< 0.05 *
Local noun Grammaticality	-0.06	0.04	-1.36	0.18
Grammaticality x *NPE*	-0.05	0.04	-1.11	0.27
Local noun x *NPE*	0.02	0.04	0.35	0.73
Local noun x *Grammaticality* x *NPE*	0.03	0.09	-0.30	0.76
Regression Path Duration				
(Intercept)	6.82	*0.08*	89.72	
Local noun	-0.19	0.09	-2.06	< 0.05 *
Grammaticality	0.14	0.10	1.43	0.16
NPE	0.52	0.11	4.97	< 0.001 ***
Local noun Grammaticality	-0.53	0.16	-3.21	< 0.01 **
Grammaticality x *NPE*	0.17	0.16	1.03	0.30
Local noun x *NPE*	0.08	0.16	0.50	0.62
Local noun x *Grammaticality* x *NPE*	-0.20	0.33	-0.61	0.54
Total Fixation Time				
(Intercept)	6.32	0.05	115.57	
Local noun	-0.02	0.05	-0.42	0.68
Grammaticality	-0.08	0.05	-1.86	0.07
NPE	0.06	0.05	1.28	0.21
Local noun Grammaticality	-0.17	0.08	-2.07	< 0.05 *
Grammaticality x *NPE*	0.01	0.08	0.15	0.88
Local noun x *NPE*	0.06	0.08	0.77	0.44
Local noun x *Grammaticality* x *NPE*	0.08	0.16	0.48	0.63
Spill-over region 2 (*the cold*)				
First Fixation Duration				
(Intercept)	5.33	0.02	222.28	
Local noun	-0.02	0.03	-0.88	0.38
Grammaticality	0.02	0.03	0.84	0.41
NPE	-0.02	0.03	-0.60	0.55
Local noun Grammaticality	-0.09	0.05	-1.68	0.10
Grammaticality x *NPE*	-0.05	0.05	-0.91	0.36
Local noun x *NPE*	0.03	0.05	0.63	0.53
Local noun x *Grammaticality* x *NPE*	0.05	0.09	0.59	0.56
Regression Path Duration				

(Intercept)	7.80	0.09	83.15	
Local noun	-0.04	0.06	-0.63	0.53
Grammaticality	0.09	0.07	1.20	0.24
NPE	-0.03	0.10	-0.29	0.77
Local noun Grammaticality	-0.30	0.13	-2.31	<0.05 *
Grammaticality x *NPE*	-0.14	0.13	-1.10	0.27
Local noun x *NPE*	0.23	0.13	1.81	0.07
Local noun x *Grammaticality* x *NPE*	0.05	0.26	0.21	0.83
Total Fixation Time				
(Intercept)	6.04	*0.08*	*78.86*	
Local noun	-0.02	0.04	-0.50	*0.62*
Grammaticality	0.09	0.04	2.10	<0.05 *
NPE	-0.13	0.06	-2.40	<0.05 *
Local noun Grammaticality	-0.11	0.09	-1.19	0.24
Grammaticality x *NPE*	-0.07	0.08	-0.92	0.36
Local noun x *NPE*	0.06	0.08	0.70	0.49
Local noun x *Grammaticality* x *NPE*	0.09	0.16	0.58	0.56

3.8.5. Discussion of Experiment 6: NPE

Experiment 6 aimed to address the nature of attraction effects in NPE in offline and online tasks. Experiment 6a 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. Experiments 6b and 6c 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 recognize the NPE-site, the parser needs to first recognize 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 recognize 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[34]. If only the head noun were retrieved, we would expect to

[34] 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

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.

In an eye-tracking while reading experiment, we observed an agreement attraction effect for both NPE and the baseline conditions. This manifested itself as an interaction between *Local noun* and *Grammaticality* in the spillover region 1 in Regression Path Duration as well as in Total Fixation Time. Further subset analysis showed that both NPE and the baseline revealed an agreement attraction where plural local nouns were read significantly faster than the singular local nouns in ungrammatical conditions. At the spillover region 2, an interaction between *Local noun* and *Grammaticality* was also observed. Note that an agreement attraction was observed at the RPD as well as at the TFT and was absent at the First Fixation Duration. Furthermore, a main effect of

---

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.

Grammaticality was observed at later stage of processing (TFT) as well as at the spillover region 2. These results can be accounted for in terms of the complex procedures the parser needs to undergo to achieve the right interpretation of the NPE; the parser needs to first recover the ellipsis site and then calculate an agreement attraction. This requires multiple processes, potentially delaying the timing of the recognition of the ungrammaticality as well as the reanalysis processes.

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[35]. 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 (Arregui et al., 2006; Callahan, Shapiro, & Love, 2010; Dickey & Bunger, 2011; Frazier, Munn, & Clifton, 2000; Kehler, 2000; Poirier, Wolfinger, Spellman, & Shapiro, 2010; Shapiro, Hestvik, Lesan, & Garcia,

---

[35] 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.

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 7 was designed to test this possibility.

3.9. Experiment 7a No Anaphora: An Acceptability Rating Experiment (Offline)

3.9.1. 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.

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 *No Anaphora* (No Anaphora vs. Baseline) were manipulated as independent factors. Experimental items were similar to those used in Experiment 6 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 summarized in Table 36. 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.

Table 36. Sample stimuli for experiment 7.

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

Factors			Examples
Local noun	Grammaticality	No Anaphora	
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.

3.9.2. Procedure

Stimuli were presented on a desktop PC using Linger software (Rohde, 2003). The task was otherwise identical to Experiment 6a.

3.9.3. Analysis

The analysis was similar to Experiment 6a[36]. Each model included simple difference sum-

---

[36] Again, residuals followed a symmetrical distribution around zero, suggesting normality (Min=-3.15; Median=-0.02; Max=3.99). As in Experiment 6a, following a reviewer's suggestion, we also carried out a cumulative logit model (also known as proportional odds model) of Experiment 7a. In this analysis, we found significant main effects of *Local noun*

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).

3.9.4. Results

Mean acceptability scores are shown in Table 37, and a summary of results is shown in Table 38. 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. 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

---

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.

interactions were observed.

The heightened effects of *local noun* and *No Anaphora* in ungrammatical conditions were supported by a subset analysis. In ungrammatical items, there were main effects of *Local noun* (*β = 0.41, SE=0.10, t=3.97, p<0.001*) and *No Anaphora* (*β = -0.30, SE=0.11, 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 (*β = -0.17, SE=0.09, t=-1.86, p=0.07*) such that items containing new non-anaphoric nouns were judged marginally less acceptable.

Table 37. Mean acceptability ratings from experiment 7a.
(Standard errors are in parentheses)

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)

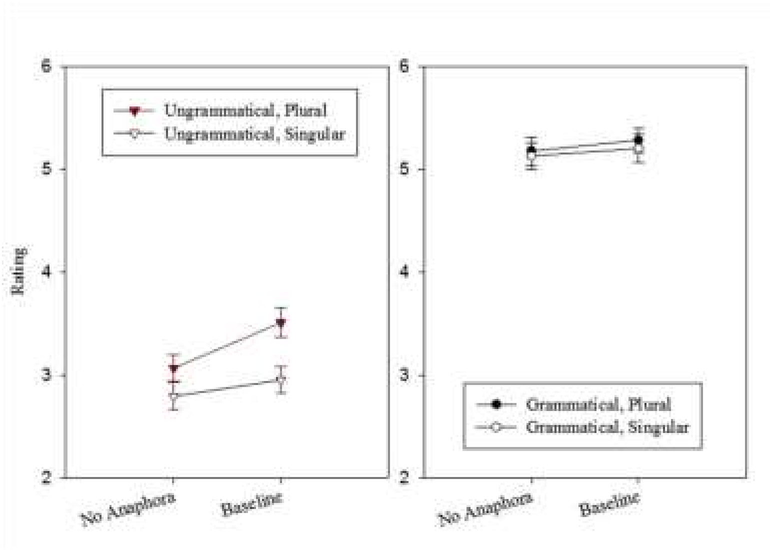


**Figure 35. Mean acceptability ratings from experiment 7a.
Error bars indicate standard error.**

Table 38. Summary of results of linear mixed effects model in experiment 7a.
Random intercepts were included for subjects and items, as were by-subject random slopes for *Local noun*, *Grammaticality*, *No Anaphora* and *Local Noun* x *Grammaticality*, and by-item random slopes for *Local noun*, *Grammaticality*, *No Anaphora*, and *Local noun* x *Grammaticality*.

	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 x *Grammaticality*	0.43	0.14	3.17	<0.001***
Grammaticality x *No Anaphora*	-0.13	0.11	-1.17	0.24
Local noun x *No Anaphora*	-0.07	0.11	-0.62	0.53
Local noun x *Grammaticality* x *No Anaphora*	-0.40	0.22	-1.84	0.07

3.10. Experiment 7b No Anaphora: A Self-Paced Word-by-Word Moving Window Experiment

3.10.1. 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 7a; items in the eight conditions were distributed in a pseudo-randomized order and combined with 74 grammatical filler sentences of similar length.

3.10.2. Procedure

The procedure was identical to Experiment 6b.

3.10.3. Analysis

Dependent measures were identical to Experiment 7a and the analysis procedure matched Experiment 6b.

3.10.4. Results

Region-by-region reading times for baseline conditions are presented in Figure 36; those for No Anaphora constructions are presented in Figure 37. Reading times at the critical spillover region for both are presented in Figure 38. Mixed effect model outputs are presented in Table 39. Mean accuracy for critical trial comprehension questions was 80.0%.

Table 39. Summary of results of linear mixed effects models by region in experiment 7b.

	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 x *Grammaticality*	-3.45	8.13	-0.42	0.67
Grammaticality x *No Anaphora*	1.27	8.13	0.16	0.88
Local noun x *No Anaphora*	11.06	8.13	1.36	0.17
Local noun x *Grammaticality* x *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 x *Grammaticality*	-11.69	9.72	-1.20	0.23
Grammaticality x *No Anaphora*	20.64	9.72	2.12	$< 0.05^{*}$
Local noun x *No Anaphora*	19.41	9.72	2.00	$< 0.05^{*}$
Local noun x *Grammaticality* x *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 x *Grammaticality*	-18.84	7.70	-2.45	$< 0.05^{*}$
Grammaticality x *No Anaphora*	4.02	7.70	0.52	0.60
Local noun x *No Anaphora*	0.07	7.70	0.01	0.99
Local noun x *Grammaticality* x *No Anaphora*	9.69	15.40	0.63	0.53

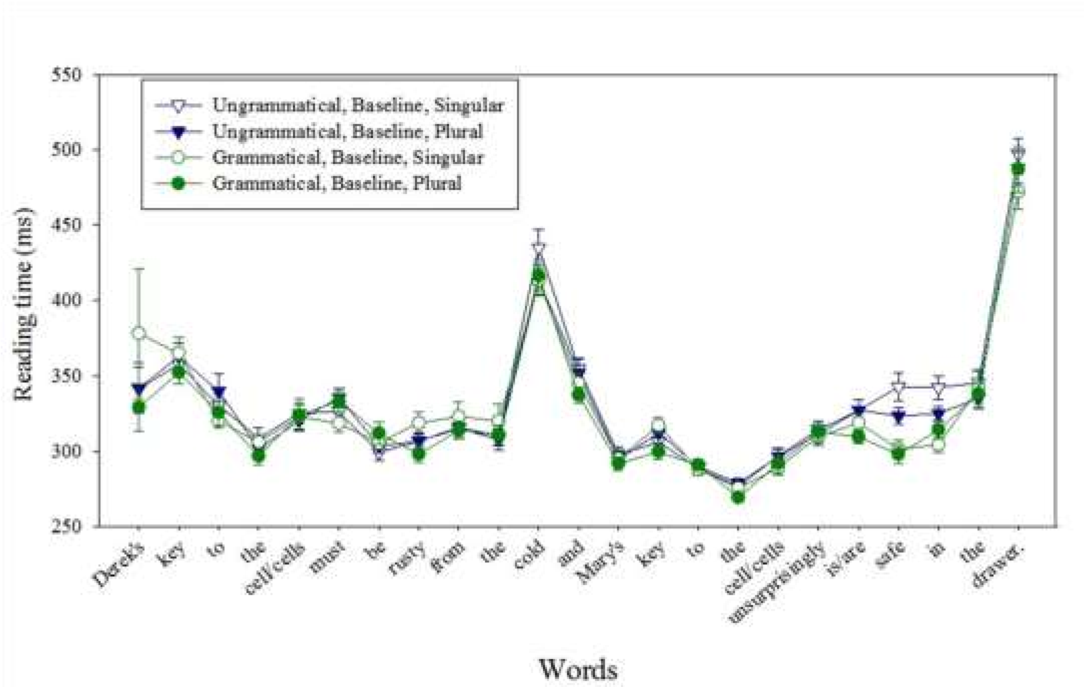


Figure 36. Region-by-region reading time means from experiment 7b for baseline conditions. Error bars indicate the standard error. The regions of interest are *is/are* (verb), *safe* (spillover 1), and *in* (spillover 2).

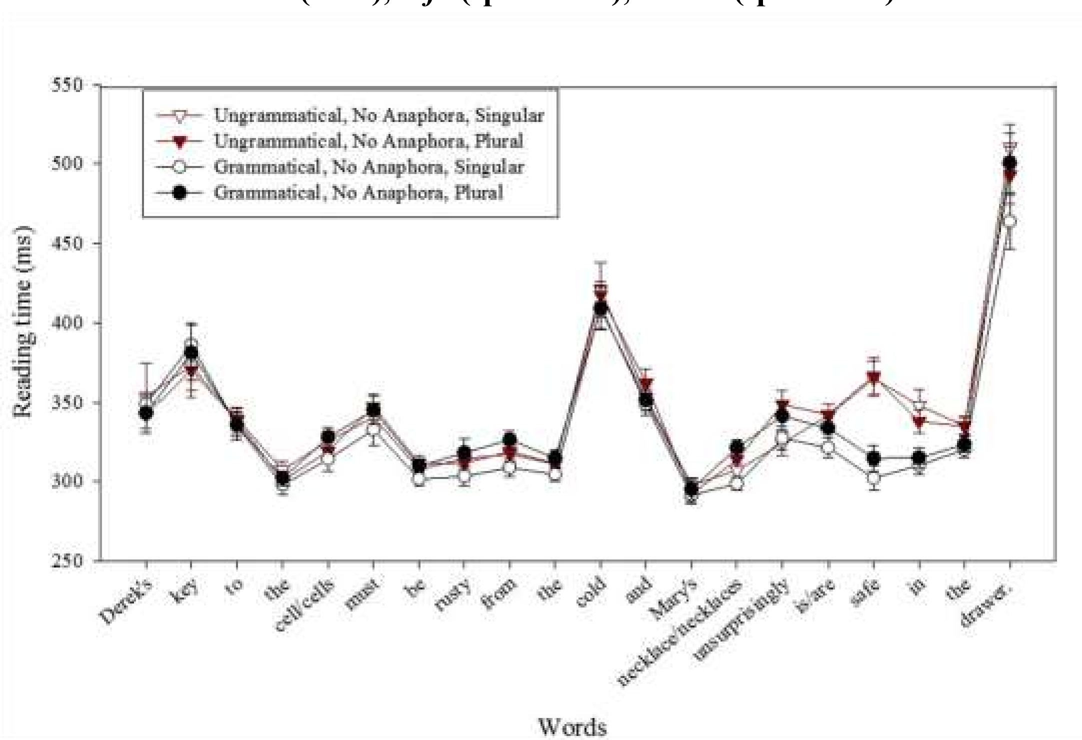


Figure 37. Region-by-region reading time means from experiment 7b for No Anaphora conditions. Error bars indicate the standard error. The regions of interest are *is/are* (verb), *safe* (spillover 1), and *in* (spillover 2).

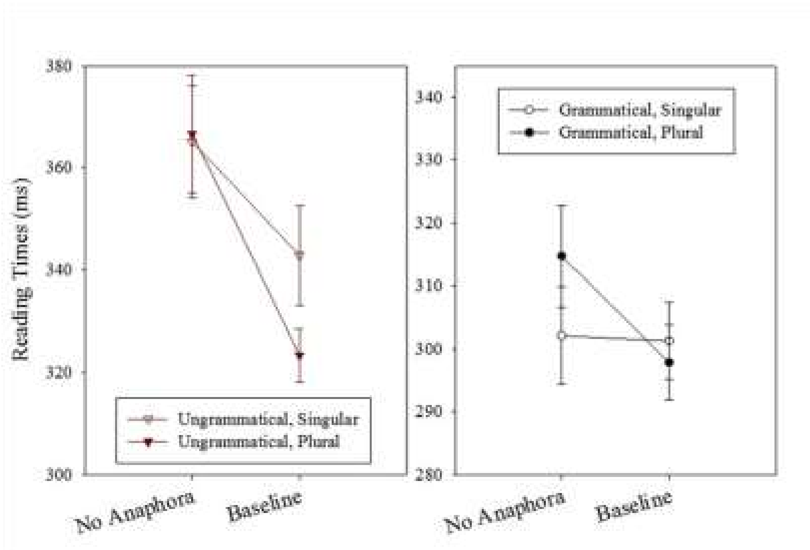


Figure 38. Reading times at the spillover region 1 (*safe*) for all conditions in experiment 7b. Error bars indicate the standard error.

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* ($\beta = 16.09, SE=9.13, t=1.76, p=0.08$) 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* ($\beta = 24.14, SE=7.56, t=3.19, p<0.01$) 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* in the No Anaphora condition ($\beta = 28.17, SE=7.71, t=3.65, p<0.001$).

3.11. Experiment 7c No Anaphora: An Eye-Tracking while Reading Experiment

3.11.1. Participants, Materials and Design

In this experiment, 60 native speakers of English from the Northwestern University community with no history of language/reading disorders participated. Before the experiment, participants provided informed consent and received course credit in an introductory Linguistics class for their participation.

Same critical items were used as in Experiment 7b. Items were distributed in a pseudo-randomized manner so that the experimental items of from the same experiment did not appear adjacent to each other. The experimental items were mixed with 72 filler sentences of similar length and complexity.

3.11.2. Procedure

Similar procedure was performed as in 6c.

3.11.3. Analysis

Similar analysis was performed as in 6c.

3.11.4. Results

<u>Verb Region</u>

In the Regression Path Duration measure, a main effect of *No Anaphora* was observed such that sentences with No Anaphora were read significantly slower than the baseline conditions. In Total Fixation Time, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower than the grammatical sentences.

<u>Spillover region 1</u>

In First Fixation Duration, a main effect of *Local noun* was observed such that items with plural local nouns were read significantly slower than those with singular local nouns. An interaction between *Grammaticality* and *No Anaphora* was also observed; further subset analysis however did not reveal any effect. In the Regression Path Duration, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower than grammatical sentences. A main effect of *No Anaphora* was observed such that sentences with No Anaphora were read significantly slower than the baseline conditions.

In the Total Fixation Time measure, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower than the grammatical sentences. A main effect of *No Anaphora* was also observed such that sentences with No Anaphora were read significantly slower than the baseline conditions. An interaction between *Local noun* and *No Anaphora* was also observed. Further subset analysis revealed only a marginal main effect of *Grammaticality* ($\beta = 0.07$ $SE = 0.04$, $t = 1.90$) in No Anaphora conditions but a main effect of *Local noun* ($\beta = -0.07$, $SE = 0.03$, $t = -2.13$, $p<0.05$), *Grammaticality* ($\beta = 0.10$, $SE = 0.03$, $t = 3.35$, $p<0.01$) and an interaction between *Local noun* and *Grammaticality* ($\beta = -0.12$, $SE = 0.07$, $t = -2.94$, $p<0.01$) in the No Anaphora conditions. Furthermore, in ungrammatical conditions, an interaction between *Local noun* and *No Anaphora* was observed ($\beta = -0.19$, $SE=0.07$, $t=-2.94$, $p<0.01$) but no effects were significant in grammatical conditions. This further suggests that plural local nouns were read significantly faster than their counterparts only in ungrammatical conditions. This was further bolstered by a three-way interaction between *Local noun*, *Grammaticality* and *No Anaphora*.

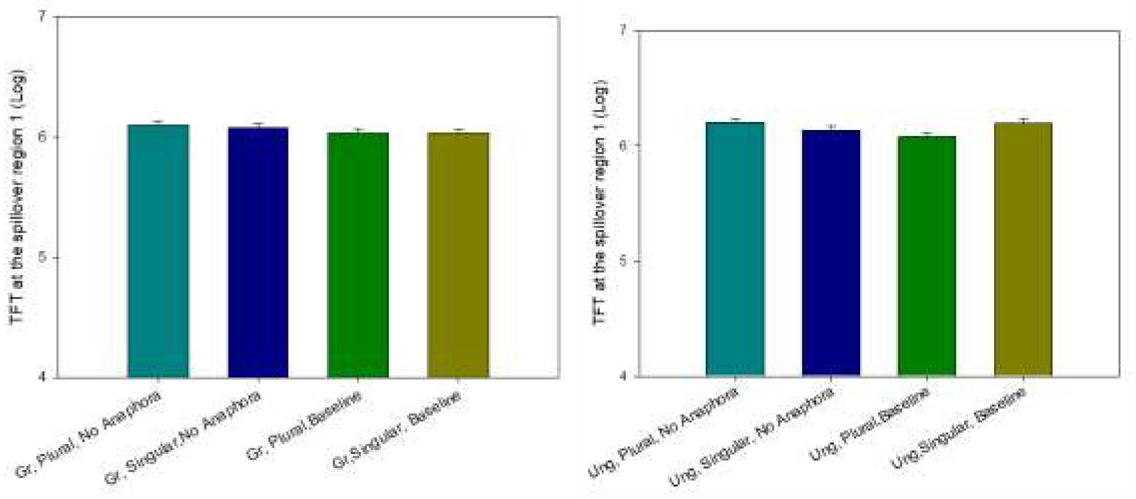


Figure 39. Bar plot for the RPD in grammatical conditions (left) and ungrammatical conditions (right) at the verb region for No Anaphora and the baseline in experiment 7c.

<u>Spillover region 2</u>

In the First Fixation Duration, an interaction between *Local noun* and *Grammaticality* was observed. However, further subset analysis did not reveal any significant effects.

Table 40. The statistical analysis of results for the eye-tracking while reading experiment (experiment 7c) on the verb region, spillover region 1, and spillover region 2.

	Estimate	SE	t	p
Verb Region (*is/are*)				
First Fixation Duration				
(Intercept)	5.45	0.03	215.67	
Local noun	0.00	*0.03*	0.17	*0.87*
Grammaticality	0.01	*0.03*	0.23	0.82
No Anaphora	-0.04	*0.03*	-1.36	0.18

Local noun x *Grammaticality*	0.05	0.04	1.08	0.28
Grammaticality x *No Anaphora*	-0.06	0.05	-1.38	0.17
Local noun x *No Anaphora*	-0.04	0.04	-0.97	0.34
Local noun x *Grammaticality* x *No Anaphora*	0.08	0.09	0.89	0.37
Regression Path Duration				
(Intercept)	5.59	0.03	169.76	
Local noun	0.01	0.03	0.37	0.71
Grammaticality	-0.00	0.04	-0.04	0.97
No Anaphora	-0.08	0.03	-2.47	< 0.05 *
Local noun x *Grammaticality*	0.02	0.06	0.30	0.76
Grammaticality x *No Anaphora*	-0.08	0.06	-1.21	0.23
Local noun x *No Anaphora*	-0.09	0.06	-1.42	0.16
Local noun x *Grammaticality* x *No Anaphora*	0.12	0.13	0.96	0.34
Total Fixation Time				
(Intercept)	5.57	0.03	211.53	
Local noun	-0.02	0.03	-0.85	0.40
Grammaticality	0.07	0.03	2.42	< 0.05 *
No Anaphora	-0.04	0.03	-1.27	0.21
Local noun x *Grammaticality*	-0.08	0.06	-1.43	0.15
Grammaticality x *No Anaphora*	0.03	0.06	0.56	0.58
Local noun x *No Anaphora*	-0.01	0.06	-0.25	0.81
Local noun x *Grammaticality* x *No Anaphora*	0.06	0.11	0.57	0.57
Spill-over region 1				
First Fixation Duration				
(Intercept)	5.52	0.02	234.43	
Local noun	0.03	0.02	2.07	< 0.05 *
Grammaticality	0.02	0.01	1.43	0.15
No Anaphora	0.02	0.02	0.99	0.33
Local noun x *Grammaticality*	0.04	0.03	1.52	0.13
Grammaticality x *No Anaphora*	0.06	0.03	2.05	< 0.05 *
Local noun x *No Anaphora*	-0.02	0.03	-0.74	0.46
Local noun x *Grammaticality* x *No Anaphora*	-0.03	0.06	-0.56	0.58
Regression Path Duration				
(Intercept)	6.07	0.06	110.25	
Local noun	0.02	0.03	0.71	0.48
Grammaticality	0.11	0.04	3.21	< 0.01 **
No Anaphora	-0.08	0.03	-2.42	< 0.05 *
Local noun x *Grammaticality*	-0.02	0.06	-0.35	0.72
Grammaticality x *No Anaphora*	-0.03	0.06	-0.45	0.65
Local noun x *No Anaphora*	-0.12	0.07	-1.70	0.10
Local noun x *Grammaticality* x *No Anaphora*	0.04	0.11	-0.40	0.69

Total Fixation Time				
(Intercept)	6.08	0.05	123.71	
Local noun	-0.02	0.02	-0.66	0.51
Grammaticality	0.09	0.02	4.22	<0.001 ***
No Anaphora	-0.05	0.02	-2.24	< 0.05 *
Local noun x *Grammaticality*	-0.03	0.04	-0.80	0.43
Grammaticality x *No Anaphora*	0.03	0.04	0.84	0.40
Local noun x *No Anaphora*	-0.11	0.04	-2.58	<0.01 **
Local noun x *Grammaticality* x *No Anaphora*	-0.17	0.08	-2.08	< 0.05 *
Spill-over region 2				
First Fixation Duration				
(Intercept)	5.43	0.03	195.37	
Local noun	0.01	0.02	0.41	0.68
Grammaticality	-0.01	0.02	-0.51	0.61
No Anaphora	0.03	0.02	1.70	0.09
Local noun x *Grammaticality*	0.07	0.03	2.34	<0.05 *
Grammaticality x *No Anaphora*	0.03	0.03	0.97	0.33
Local noun x *No Anaphora*	-0.04	0.03	-1.33	0.18
Local noun x *Grammaticality* x *No Anaphora*	0.01	0.06	0.19	0.85
Regression Path Duration				
(Intercept)	7.09	0.08	85.27	
Local noun	0.01	0.04	0.21	*0.83*
Grammaticality	0.05	0.03	1.37	0.17
No Anaphora	0.05	0.04	1.35	0.18
Local noun x *Grammaticality*	0.00	0.07	0.07	0.95
Grammaticality x *No Anaphora*	0.07	0.07	1.01	0.31
Local noun x *No Anaphora*	-0.12	0.07	-1.74	0.08
Local noun x *Grammaticality* x *No Anaphora*	-0.00	0.14	-0.02	0.98
Total Fixation Time				
(Intercept)	5.95	0.06	100.47	
Local noun	0.03	0.03	1.29	*0.21*
Grammaticality	-0.02	0.02	-0.81	0.42
No Anaphora	0.01	0.03	0.25	0.80
Local noun x *Grammaticality*	0.08	0.05	1.70	0.09
Grammaticality x *No Anaphora*	0.06	0.05	1.23	0.22
Local noun x *No Anaphora*	-0.05	0.05	-1.01	0.31
Local noun x *Grammaticality* x *No Anaphora*	0.12	0.09	1.35	0.18

3.11.5. Discussion of Experiment 7: No Anaphora

The goal of Experiments 7 was to rule out the possibility that agreement attraction in NPE is due to coordination alone. This was done by replacing the 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 7a 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 6a. 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 7b and 7c 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 7b, 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 6b, 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 & Clifton, 2001; Frazier et al., 2000; 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 7b contrasts with the offline judgment task presented in Experiment 7a, 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.

Finally, in an eye-tracking while reading experiment (experiment 7c), we observed no agreement attraction for No Anaphora condition in none of the measures or regions. Further subset analysis showed agreement attraction in the ungrammatical baseline conditions only. We observed a main effect of *Grammaticality* at the earlier stage of processing (verb region) in Total Fixation Time, and at the spillover region 1 both in Total Fixation Time and in Regression Path Duration. This suggests that, contrary to the resolution of NPE, which involves the inspection of agreement relation and the recovery of the content of the NPE-site, No Anaphora conditions do not require the recovery of the ellipsis site. This would contribute to the earlier manifestation of the main effect of *Grammaticality* for No Anaphora condition.

3.12. Experiment 8a Anaphoric one: An Acceptability Rating Experiment (Offline)

3.12.1. Participants & Materials and Design

Participants were 52 native speakers of English from the Northwestern University community with no history of 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 summarized in Table 41. Items were similar to Experiments 6 and 7, 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.

Table 41. Sample stimuli for experiment 8.

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

Factors			
Local noun	Grammaticality	Anaphoric *one*	Examples
Plural	*Grammatical*	Anaphoric *one*	…Mary's dull one unsurprisingly is safe in the drawer.
Plural	*Ungrammatical*	Anaphoric *one*	…Mary's dull one unsurprisingly are safe in the drawer.
Singular	*Grammatical*	Anaphoric *one*	…Mary's dull one unsurprisingly is safe in the drawer.
Singular	*Ungrammatical*	Anaphoric *one*	…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.

3.12.2. Procedure

The procedure was the same as Experiment 7a.

3.12.3. Analysis

Analysis was similar to Experiment 6a and 7a; 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[37].

---

[37] Residuals followed a symmetrical distribution around zero, suggesting normality (Min=-4.53; Median=0.06; Max=3.12). As in Experiments 6a and 7a, we also carried out a cumulative logit model (also known as proportional odds model) of Experiment 8a. This

3.12.4. Results

Mean acceptability scores are shown in Table 42, and mixed effect model outputs are shown in Table 43. A marginal main effect of *Local noun* was observed such 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.

---

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* ($\beta = -0.65, SE= 0.34, z= -1.92, p=0.06$).

Table 42. Mean acceptability ratings from experiment 8a.
(Standard errors are in parentheses)

Factors			
Local noun	Grammaticality	Anaphoric *one*	Average raw rating (SE)
Plural	*Grammatical*	Anaphoric *one*	5.36 (0.12)
Plural	*Ungrammatical*	Anaphoric *one*	3.91 (0.12)
Singular	*Grammatical*	Anaphoric *one*	5.38 (0.13)
Singular	*Ungrammatical*	Anaphoric *one*	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)

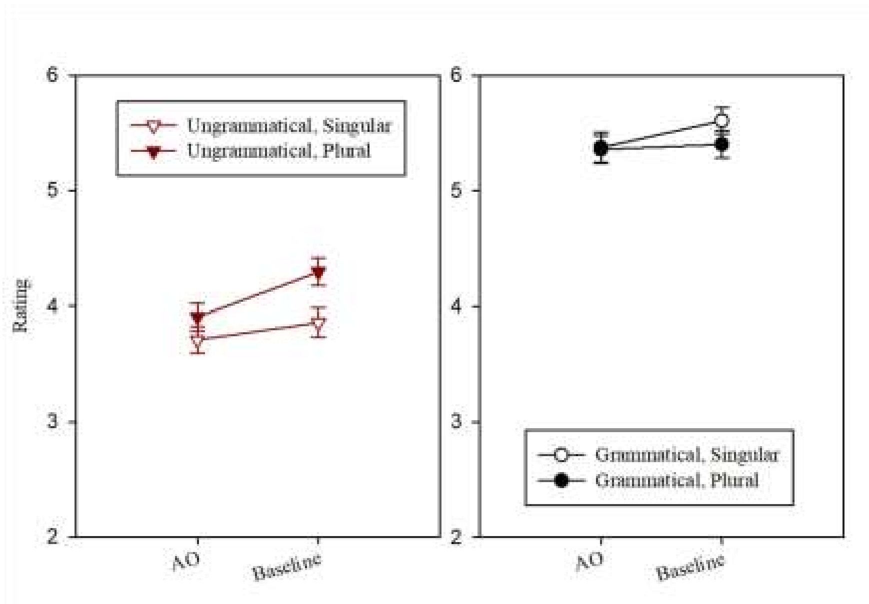


Figure 40. Mean acceptability ratings from experiment 8a.
Error bars indicate standard error.

Table 43. Summary of results of linear mixed effects models in experiment 8a.
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*.

	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 x *Anaphoric one*	-0.10	0.11	-0.92	0.36
Local noun x *Anaphoric one*	-0.05	0.11	-0.48	0.63
Local noun x *Grammaticality* x *Anaphoric one*	-0.51	0.21	-2.40	<0.05*

3.13. Experiment 8b Anaphoric one: A Self-Paced Word-by-Word Moving Window Experiment

3.13.1. 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 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 8a (see Table 41), but an adjective was included in the first conjunct to increase the diversity and naturalness of the items. To ensure that participants did not encounter the same types of target items consecutively, 32 items were distributed in a pseudo-randomized manner. In addition to the current experimental items, there were 74 filler sentences that involved irrelevant manipulations to the current ones.

3.13.2. Procedure

The procedure was identical to Experiment 6b and 7b.

3.13.3. Analysis

Dependent measures were the same as Experiment 8a, and the analysis procedure matched

Experiment 6b and Experiment 7b.

3.13.4. Results

The region-by-region reading times for baseline conditions are presented in Figure 41; those for Anaphoric *one* (AO) constructions are presented in Figure 42. Reading times at the critical spillover region for both are presented in Figure 43. Mixed effect model outputs are presented in Table 44. Mean accuracy for critical trial comprehension questions was 75%.

Table 44. Summary of results of linear mixed effects models by region in experiment 8b.

	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*
Anaphoric *one*	11.99	4.42	2.71	<0.01**
Local noun x *Grammaticality*	-5.22	7.11	-0.73	0.46
Grammaticality x Anaphoric *one*	2.68	7.11	0.38	0.71
Local noun x Anaphoric *one*	2.69	7.11	0.38	0.71
Local noun x *Grammaticality* x Anaphoric *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* x Grammaticality, and *Local noun* x 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***
Anaphoric one	12.28	4.35	2.82	<0.01**
Local noun x *Grammaticality*	3.95	9.19	0.43	0.67
Grammaticality x Anaphoric *one*	12.64	6.80	1.86	0.06
Local noun x Anaphoric *one*	9.31	7.63	1.22	0.22
Local noun x *Grammaticality* x Anaphoric *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* x *Anaphoric one*, and *Local noun* x *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**
Anaphoric *one*	2.71	4.13	0.66	0.51
Local noun x *Grammaticality*	8.69	6.22	1.40	0.16
Grammaticality x Anaphoric *one*	6.36	7.57	0.84	0.40
Local noun x Anaphoric *one*	1.30	6.65	0.20	0.84
Local noun x *Grammaticality* x Anaphoric *one*	16.70	12.44	1.34	0.18

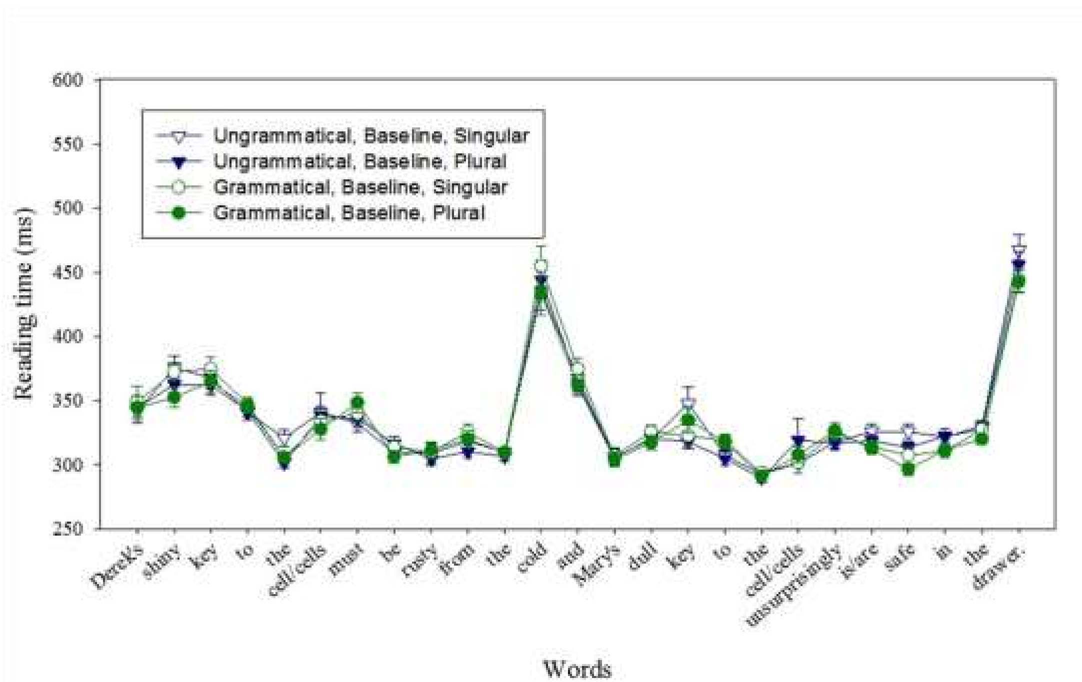


Figure 41. Region-by-region reading time means from experiment 8b for baseline conditions. Error bars indicate the standard error. The regions of interest are *is/are* (verb), *safe* (spillover 1), and *in* (spillover 2).

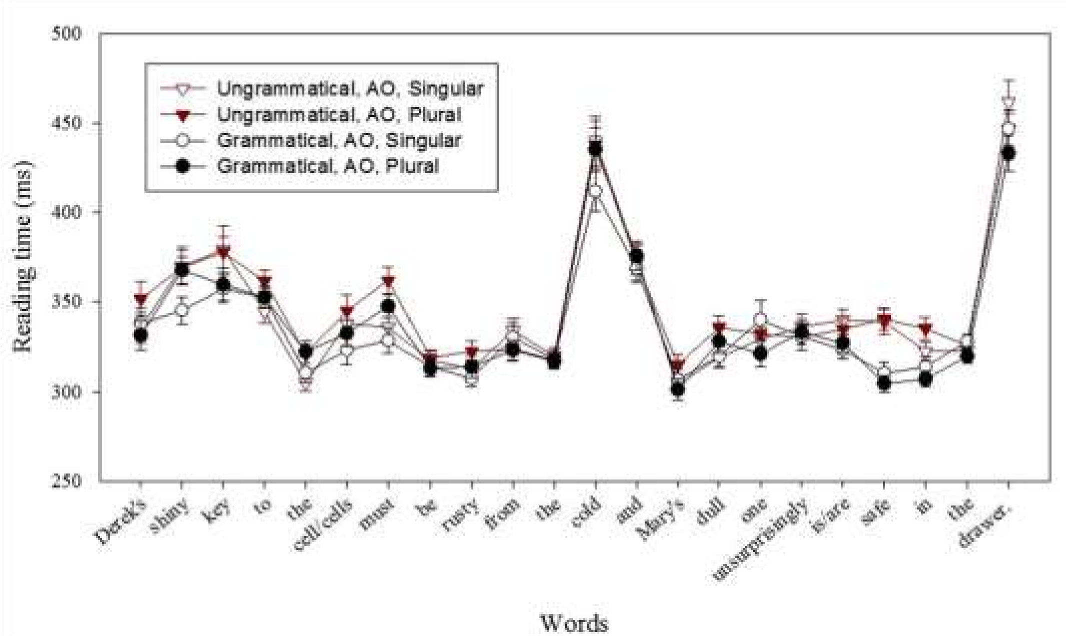


Figure 42. Region-by-region reading time means from experiment 8b for Anaphoric *one* conditions. Error bars indicate the standard error. The regions of interest are *is/are* (verb), *safe* (spillover 1), and *in* (spillover 2).

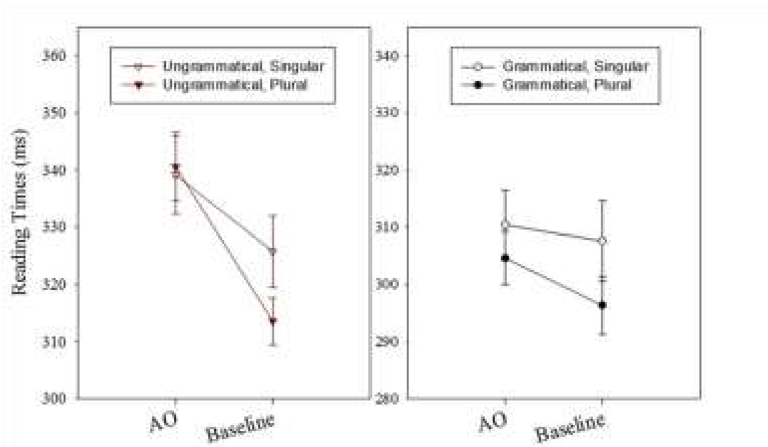


Figure 43. Reading times at the spillover region (*safe*) for all conditions in experiment 8b. Error bars indicate the standard error.

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 6) and Anaphoric *one* (Experiment 8) were directly compared.[38] At the spill-over region 1, there was a significant three-way interaction between *Construction type* (NPE/Anaphoric *one*), *Grammaticality* and a *Local noun* ($\beta = -34.38$, $SE=15.92$, $t=-2.16$, $p<0.05$). 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.

---

[38] This was suggested by an anonymous reviewer from *Language, Cognition, and Neuroscience*. We are grateful to the reviewer for pointing out this possibility.

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.

3.14. Experiment 8c Anaphoric one: An Eye-Tracking while Reading Experiment

3.14.1. Participants, Materials and Design

In this experiment, 52 native speakers of English from Northwestern University with no history of language/reading disorders participated. Before the experiment, participants provided informed consent and received course credit in an introductory Linguistics class for their participation.

The same critical items were used as in Experiment 8b. Items were distributed in a pseudo-randomized manner so that the experimental items from the same experiment did not appear adjacent to each other. The experimental items were mixed with 80 filler sentences of similar length and complexity.

3.14.2. Procedure

Similar procedure was performed as in 6c.

3.14.3. Analysis

Similar analysis was performed as in 6c.

3.14.4. Results

Verb Region

In First Fixation Duration, we observed a main effect of Anaphoric *one* such that sentences with Anaphoric *one* were read significantly slower than the baseline conditions. In Regression Path Duration, a main effect of Anaphoric *one* was observed such that the baseline conditions were read significantly slower than the constructions involving Anaphoric *one*. This was driven by the interaction between *Local noun* and Anaphoric *one*. Further subset analysis showed a main effect of *Local noun* in the Anaphoric *one* conditions ($\beta = 0.36, SE= 0.16, t= 2.29, p<0.05$). In the Total Fixation Time measure, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower than their ungrammatical counterparts.

Spillover Region 1

In the Regression Path Duration, a main effect of *Local noun* was observed. Further subset analysis revealed a main effect of *Local noun* ($\beta = -0.16, SE= 0.15, t= -3.46, p<0.01$) and a marginal interaction between *Local noun* and *Grammaticality* ($\beta = -0.14, SE= 0.09\ t= -1.67$) in the baseline condition, but not in the Anaphoric *one* conditions. In the Total Fixation Time measure, we also observed a main effect of *Local noun*. For all measures, there was no interaction between *Local noun* and *Grammaticality*.

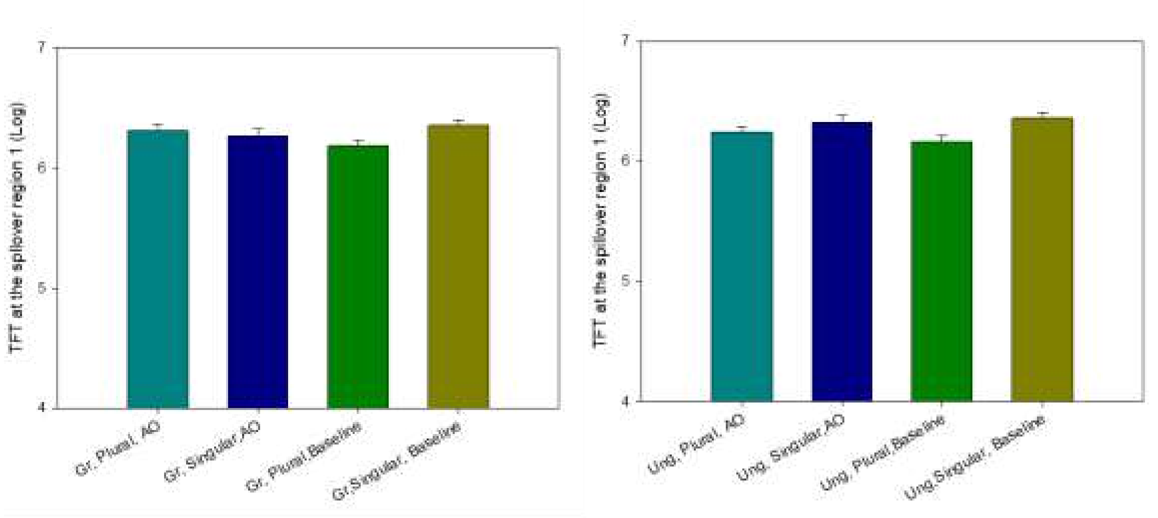


Figure 44. Bar plot for the TFT in grammatical conditions (left) and ungrammatical conditions (right) at the spillover region 1 for Anaphoric *one* and the baseline in experiment 8c.

<u>Spillover Region 2</u>

In Regression Path Duration, a main effect of *Local noun* was observed such that singular local nouns were read significantly slower than the plural local nouns. A main effect of Anaphoric *one* was observed such that sentences with Anaphoric *one* were read significantly slower than the baseline conditions. In the Total Fixation Time, a main effect of Anaphoric *one* was also observed. An interaction between *Local noun* and Anaphoric *one* was also observed. No other effects reached significance.

Table 45. The statistical analysis of results for the eye-tracking while reading experiment (experiment 8c) on the verb region, spillover region 1, and spillover region 2.

	Estimate	SE	t	p
Verb Region (*is/are*)				
First Fixation Duration				
(Intercept)	5.32	0.04	149.52	
Local noun	0.09	0.07	1.35	*0.18*
Grammaticality	-0.04	0.06	-0.72	0.49

Anaphoric *one*	-0.16	0.06	-2.67	<0.05 *
Local noun x *Grammaticality*	-0.03	0.11	-0.31	0.77
Grammaticality x Anaphoric *one*	-0.08	0.10	-0.79	0.44
Local noun x Anaphoric *one*	-0.13	0.11	-1.19	0.25
Local noun x *Grammaticality* x Anaphoric *one*	-0.04	0.20	-0.20	0.85
Regression Path Duration				
(Intercept)	6.07	*0.08*	*74.80*	
Local noun	0.05	0.12	0.40	0.69
Grammaticality	0.18	0.14	1.33	0.20
Anaphoric *one*	0.55	0.17	3.15	<0.01 **
Local noun x *Grammaticality*	-0.09	0.24	-0.38	0.70
Grammaticality x Anaphoric *one*	0.01	0.24	0.03	0.98
Local noun x Anaphoric *one*	-0.58	0.24	-2.47	<0.05 *
Local noun x *Grammaticality* x Anaphoric *one*	0.60	0.48	1.25	0.22
Total Fixation Time				
(Intercept)	*5.62*	*0.38*	*147.85*	
Local noun	0.09	0.07	1.30	0.20
Grammaticality	0.27	0.07	4.04	<0.001 ***
Anaphoric *one*	0.01	0.07	0.13	0.90
Local noun x *Grammaticality*	-0.10	0.12	-0.81	0.42
Grammaticality x Anaphoric *one*	0.03	0.12	0.29	0.78
Local noun x Anaphoric *one*	-0.17	0.12	-1.39	0.18
Local noun x *Grammaticality* x Anaphoric *one*	-0.19	0.24	-0.80	0.44
Spill-over region 1				
First Fixation Duration				
(Intercept)	5.43	0.02	358.9	
Local noun	0.02	0.02	0.7	0.48
Grammaticality	-0.04	0.02	-1.8	0.08
Anaphoric *one*	0.01	0.03	0.5	0.59
Local noun x *Grammaticality*	0.06	0.04	1.4	0.16
Grammaticality x Anaphoric *one*	0.04	0.04	1.1	0.29
Local noun x Anaphoric *one*	-0.02	0.04	-0.5	0.60
Local noun x *Grammaticality* x Anaphoric *one*	-0.16	0.08	-1.9	0,05
Regression Path Duration				
(Intercept)	*6.66*	*0.07*	90.30	
Local noun	-0.25	0.07	-3.57	<0.001 ***
Grammaticality	0.11	0.06	1.69	0.09
Anaphoric *one*	-0.07	0.09	-0.80	0.42
Local noun x *Grammaticality*	-0.10	0.13	-0.79	0.43
Grammaticality x Anaphoric *one*	-0.11	0.13	-0.84	0.4
Local noun x Anaphoric *one*	-0.09	0.123	-0.70	0.49

Local noun x *Grammaticality* x Anaphoric *one*	-0.05	0.26	-0.20	0.84
Total Fixation Time				
(Intercept)	*6.25*	*0.05*	*122.17*	
Local noun	-0.08	0.04	-2.27	<0.05 *
Grammaticality	-0.01	0.03	-0.34	0.73
Anaphoric *one*	-0.04	0.04	-1.02	0.32
Local noun x *Grammaticality*	-0.10	0.06	-1.68	0.09
Grammaticality x Anaphoric *one*	0.04	0.06	0.70	0.49
Local noun x Anaphoric *one*	-0.14	0.06	-2.23	<0.05 *
Local noun x *Grammaticality* x Anaphoric *one*	-0.03	0.12	-0.28	0.78
Spill-over region 2				
First Fixation Duration				
(Intercept)	5.37	0.02	279.75	
Local noun	0.01	0.02	0.25	0.80
Grammaticality	-0.01	0.02	-0.23	0.82
Anaphoric one	-0.03	0.02	-1.52	0.14
Local noun x *Grammaticality*	0.04	0.04	0.88	0.38
Grammaticality x *Anaphoric one*	-0.06	0.04	-1.31	0.19
Local noun x *Anaphoric one*	0.02	0.04	0.49	0.62
Local noun x *Grammaticality* x *Anaphoric one*	-0.00	0.09	-0.04	0.97
Regression Path Duration				
(Intercept)	7.54	*0.09*	*79.71*	
Local noun	-0.12	0.05	-2.19	<0.05 *
Grammaticality	-0.08	0.06	-1.38	0.17
Anaphoric *one*	-0.34	0.08	-4.23	<0.001 ***
Local noun x *Grammaticality*	-0.02	0.11	-0.20	0.84
Grammaticality x Anaphoric *one*	-0.01	0.11	-0.11	0.91
Local noun x Anaphoric *one*	-0.11	0.11	-1.00	0.32
Local noun x *Grammaticality* x Anaphoric *one*	0.04	0.21	0.18	0.85
Total Fixation Time				
(Intercept)	5.97	0.06	96.29	
Local noun	-0.01	0.04	-0.32	0.75
Grammaticality	0.01	0.03	0.34	0.74
Anaphoric *one*	-0.17	0.04	-4.30	<0.001 ***
Local noun x *Grammaticality*	-0.04	0.07	-0.61	0.55
Grammaticality x Anaphoric *one*	-0.00	0.06	-0.06	0.95
Local noun x Anaphoric *one*	-0.15	0.06	-2.35	<0.05 *
Local noun x *Grammaticality* x Anaphoric *one*	0.03	0.13	0.25	0.81

3.14.5. Discussion of Experiment 8: Anaphoric one

Experiment 8 aimed to further investigate how the parsers' sensitivity to grammatical distinction impacts processing of anaphoric elements, replacing the NPE in Experiment 6 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 8a) 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 8b & 8c) 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 recognized, the parser can access and reactivate the antecedent of Anaphoric *one*.

(73) 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.[39] 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.

(74) ... and Mary's dull **one** unsurprisingly are ...

Once the agreement mismatch is recognized, 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.

---

[39] 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.

(75) 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 8 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.

The results of the eye-tracking experiment of Anaphoric *one* showed no agreement attraction for Anaphoric *one*, as revealed by an absence of an interaction between *Local noun* and *Grammaticality* in all of the regions and measures. This further supports that the retrieved information for NPE and Anaphoric *one* is different. Anaphoric *one* provides a strong morphological cue that marks the singular noun, and therefore the reader should easily retrieve the head noun, making it unnecessary to retrieve the local noun. That is, there are multiple aspects of the head that match the retrieval cues associated with Anaphoric *one*, such as a singular NP and noun category. Therefore, the parser may possibly select the head in the antecedent as the feasible subject. Also note that a main effect of *Grammaticality* was observed at the verb region, which is in sharp contrast with

NPE. The absence of the agreement attraction and the manifestation of grammaticality at the early stage for Anaphoric *one* may also be due to its nature as a deep anaphora where it requires semantics and discourse related information of the antecedent rather than the retrieval of grammatical information associated with an antecedent.

3.15. Experiment 9a Pronoun it: An Acceptability Rating Experiment (Offline)

3.15.1. Participants & Materials and Design

For this experiment, 39 native English speakers from Northwestern University with no history of language disorders participated and gave informed consent. In exchange for their participation, the participants were granted 1 credit for introductory linguistic classes taught at Northwestern.

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 Pronoun *it* (Pronoun *it* vs. baseline) were manipulated as independent factors. A sample set of stimuli is summarized in Table 46. Items were similar to Experiments 6 and 8, but contained items with Pronoun *it* rather than NPE or Anaphoric *one* condition. To ensure that participants did not encounter the same types of target items consecutively, 32 items were distributed in a pseudo-randomized manner. In addition to the current experimental items, there were 74 filler sentences that involved irrelevant manipulations to the current ones.

Table 46. Sample stimuli for experiment 9.
Derek's key to the cell/cells must be rusty from the cold and...

Factors			Examples
Local noun	Grammaticality	Pronoun *it*	
Plural	*Grammatical*	*Pronoun it*	… it unsurprisingly is stuck in the drawer.
Plural	*Ungrammatical*	*Pronoun it*	… it unsurprisingly are stuck in the drawer.
Singular	*Grammatical*	*Pronoun it*	… it unsurprisingly is stuck in the drawer.
Singular	*Ungrammatical*	*Pronoun it*	… it unsurprisingly are stuck in the drawer.
Plural	*Grammatical*	*Baseline*	… Mary's key to the cells unsurprisingly is stuck in the drawer.
Plural	*Ungrammatical*	*Baseline*	… Mary's key to the cells unsurprisingly are stuck in the drawer.
Singular	*Grammatical*	*Baseline*	… Mary's key to the cell unsurprisingly is stuck in the drawer.
Singular	*Ungrammatical*	*Baseline*	… Mary's key to the cell unsurprisingly are stuck in the drawer.

3.15.2. Procedure

The procedure was the same as Experiment 8a.

3.15.3. Analysis

Analysis was similar to Experiment 8a; fixed effects were *Local noun number* (singular vs. plural), *Grammaticality* (grammatical vs. ungrammatical), and Pronoun *it* (whether the sentences involved *it* vs baseline) and their interactions.

3.15.4. Results

Mean acceptability scores are shown in Table 47, and mixed effect model outputs are shown in Table 48. A main effect of *Grammaticality* was observed such that ungrammatical sentences were rated less acceptable than grammatical constructions. A main effect of Pronoun *it* was also observed such that sentences containing Pronoun *it* were rated less acceptable. These effects were driven by an interaction between *Local noun* and *Grammaticality* as ungrammatical singular items were rated less acceptable compared to their counterparts. An interaction between *Grammaticality* and Pronoun *it* was also observed as well as an interaction between *Local noun* and Pronoun *it* was observed: in ungrammatical conditions, a main effect of *Local noun* ($\beta = 0.28, SE=0.11, t=-2.46, p<0.05$) as well as a main effect of Pronoun *it* ($\beta = -0.46, SE=0.12, t=-3.80, p<0.001$) was observed, but not in grammatical conditions

Table 47. Mean acceptability ratings from experiment 9a.

(Standard errors are in parentheses)

Factors			Average raw rating (SE)
Local noun	Grammaticality	Pronoun *it*	
Plural	*Grammatical*	Pronoun *it*	5.12 (0.13)
Plural	*Ungrammatical*	Pronoun *it*	3.37 (0.13)
Singular	*Grammatical*	Pronoun *it*	5.23 (0.13)
Singular	*Ungrammatical*	Pronoun *it*	3.24 (0.11)
Plural	*Grammatical*	*Baseline*	5.30 (0.11)
Plural	*Ungrammatical*	*Baseline*	3.97 (0.13)
Singular	*Grammatical*	*Baseline*	5.17(0.13)
Singular	*Ungrammatical*	*Baseline*	3.55 (0.12)

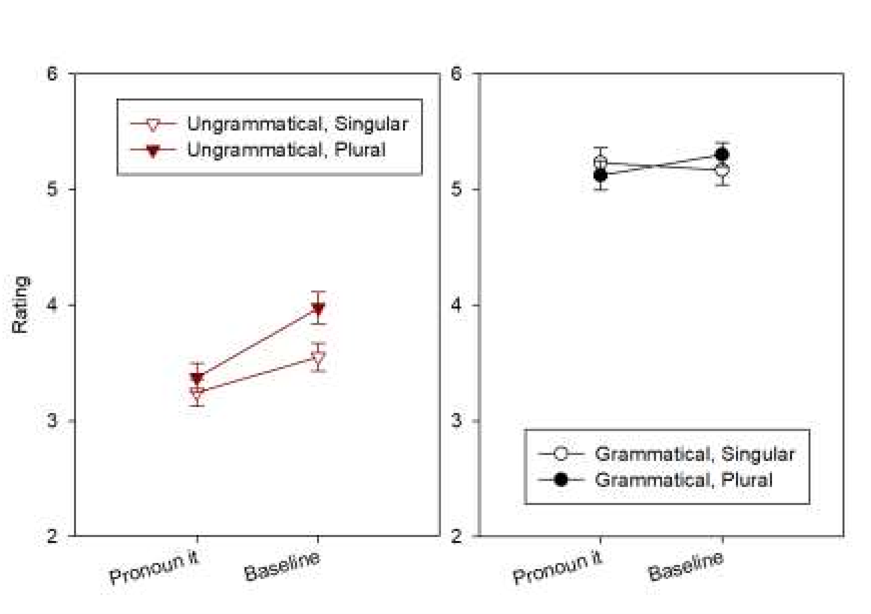


Figure 45. Mean acceptability ratings from Experiment 9a.
Error bars indicate standard error.

Table 48. Summary of results of linear mixed effects models in experiment 9a. Random intercepts were included for subjects and items, as were by-subject random slopes for *Grammaticality*, and *Pronoun it*, and by-item random slopes for *Local noun*, *Grammaticality* and *Pronoun it*.

	Estimate	SE	t	p
(Intercept)	*4.37*	*0.13*	*33.35*	
Local noun	0.15	0.08	1.89	0.07
Grammaticality	-1.67	0.20	-8.22	<0.001 ***
Pronoun it	-0.26	0.08	-3.04	< 0.01 **
Local noun x *Grammaticality*	0.26	0.13	2.00	<0.05 *
Grammaticality x *Pronoun it*	-0.40	0.13	-3.05	<0.01 **
Local noun x *Pronoun it*	-0.27	0.13	-2.07	< 0.05 *
Local noun x *Grammaticality* x *Pronoun it*	-0.05	0.26	-0.20	0.84

3.16. Experiment 9b Pronoun it: A Self-Paced Word-by-Word Moving Window Experiment

3.16.1. Participants & Materials and Design

For this experiment, 83 native English speakers from Northwestern University with no history of language disorders participated and gave informed consent. In exchange for their participation, the participants were granted 1 credit for introductory linguistic classes taught at Northwestern.

One participant was excluded because the participant's comprehension question accuracy rate was below 60%. Similar critical items were used as in Experiment 9a (see Table 46). To ensure that participants did not encounter the same types of target items consecutively, 32 items were distributed in a pseudo-randomized manner. In addition to the current experimental items, there were 74 filler sentences that involved irrelevant manipulations to the current ones.

3.16.2. Procedure

The procedure was identical to Experiment 6b.

3.16.3. Analysis

Dependent measures were the same as Experiment 9a, and the analysis procedure matched Experiment 6b and Experiment 7b.

3.16.4. Results

The region-by-region reading times for baseline conditions are presented in Figure 46; those for Pronoun *it* (*it*) constructions are presented in Figure 47. Reading times at the critical spillover region for both are presented in Figure 48. Mixed effect model outputs are presented in Table 49. Mean accuracy for critical trial comprehension questions was 76%.

Table 49. Summary of results of linear mixed effects models by region in experiment 9b.

	Estimate	SE	t	p
Verb Region (*is/are*) by-subject random intercepts and slopes for *Pronoun it* and *Local noun*, and by-item random intercepts and slopes for *Local noun* and *Grammaticality*.				
(Intercept)	303.72	7.04	43.16	
Local noun	2.99	2.95	1.01	*0.31*
Grammaticality	5.31	2.72	1.95	0.05
Pronoun it	-10.76	3.26	-3.30	< 0.01 **
Local noun x *Grammaticality*	-5.27	5.32	-0.99	0.32
Grammaticality x *Pronoun it*	-2.33	5.32	-0.44	0.66
Local noun x *Pronoun it*	1.07	5.32	0.20	0.84
Local noun x *Grammaticality* x *Pronoun it*	-5.80	10.64	-0.55	0.59
Verb Region Spill-Over Region (*stuck*) by-subject random intercepts and slopes for *Local noun*, *Grammaticality*, *Pronoun it,* and by-item random intercepts and slopes for *Local noun* and *Grammaticality*				
(Intercept)	303.44	7.19	42.19	
Local noun	-4.19	2.88	-1.45	0.15
Grammaticality	18.97	3.39	5.60	<0.001***
Pronoun it	1.78	2.81	0.63	0.53
Local noun x *Grammaticality*	-10.90	5.37	-2.03	< 0.05 *
Grammaticality x *Pronoun it*	-0.07	5.37	-0.01	0.99
Local noun x *Pronoun it*	-2.36	5.37	-0.44	0.66
Local noun x *Grammaticality* x *Pronoun it*	11.79	10.73	1.10	0.27
Verb Spill-over Region 2 (*in*) by-subject random slopes, and by-subject intercepts for *Local noun*, and *Pronoun it*, and by-item random slopes, and by-item intercepts for *Pronoun it*				
(Intercept)	304.83	6.44	47.37	
Local noun	-2.51	2.67	-0.93	0.35
Grammaticality	7.11	2.56	2.77	<0.01**
Pronoun it	-1.44	3.12	-0.45	0.65

Local noun x *Grammaticality*	-7.95	5.12	-1.55	0.12
Grammaticality x *Pronoun it*	6.74	5.12	1.32	0.19
Local noun x *Pronoun it*	5.23	5.12	1.03	0.30
Local noun x *Grammaticality* x *Pronoun it*	-5.35	10.23	-0.52	0.60

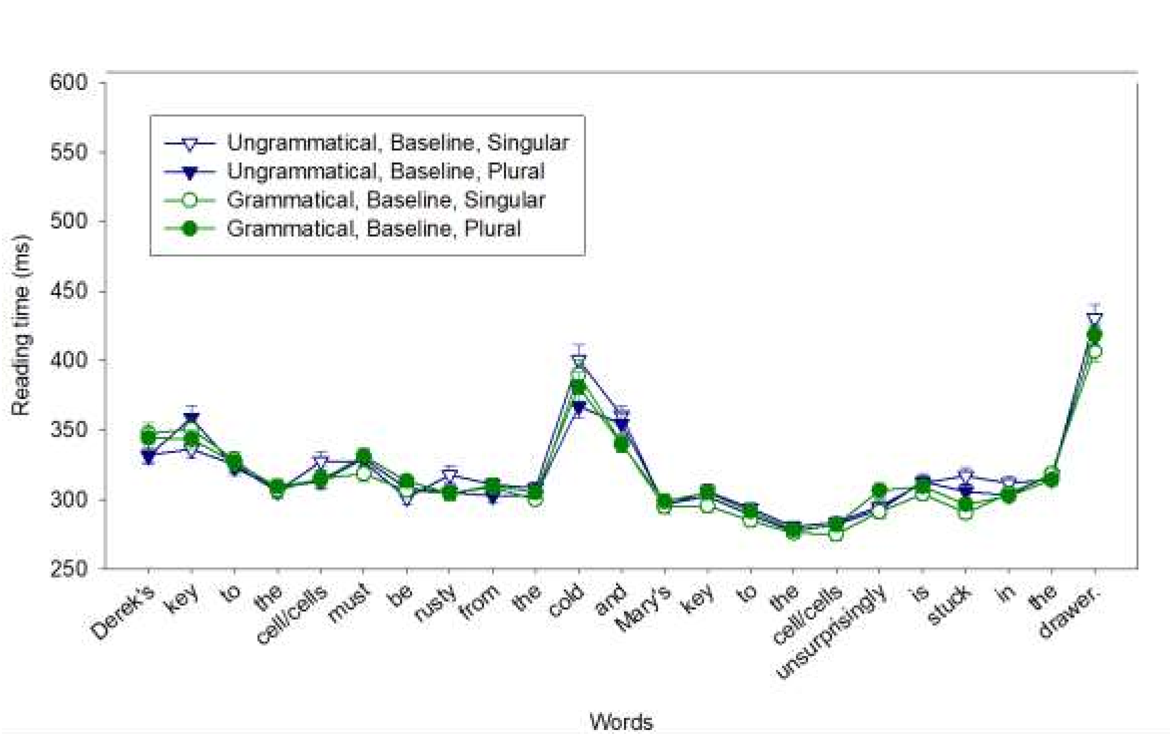


Figure 46. Region-by-region reading time means from experiment 9b for baseline conditions. Error bars indicate the standard error. The regions of interest are *is/are* (verb), *safe* (spillover 1), and *in* (spillover 2).

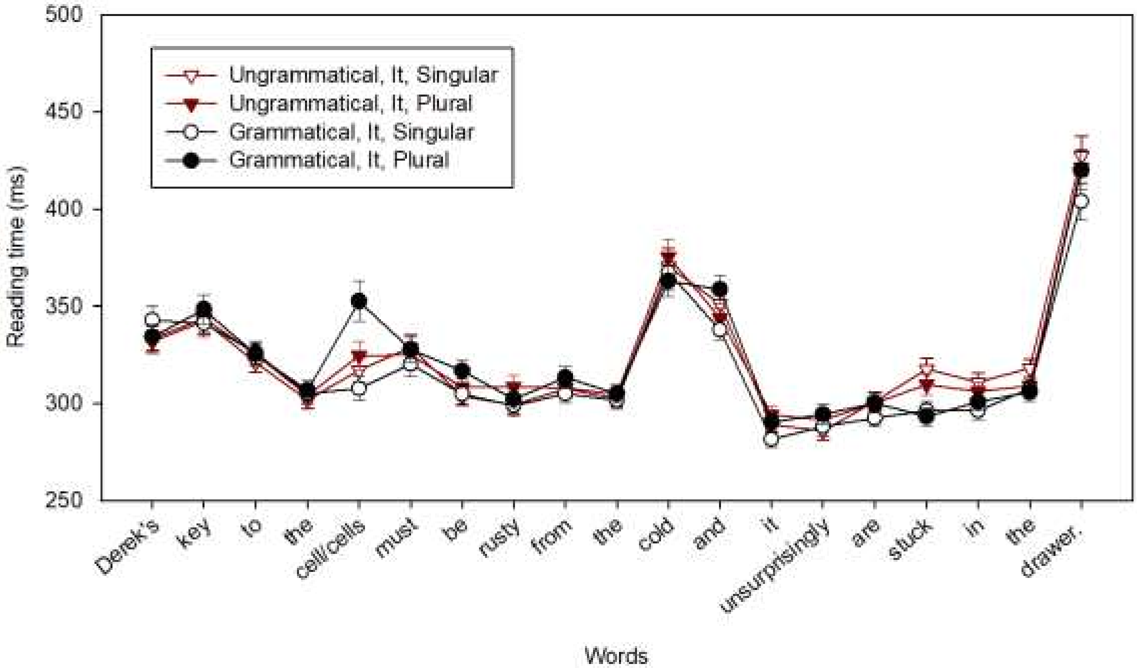


Figure 47. Region-by-region reading time means from experiment 9b for Pronoun *it* conditions. Error bars indicate the standard error. The regions of interest are *is/are* (verb), *safe* (spillover 1), and *in* (spillover 2).

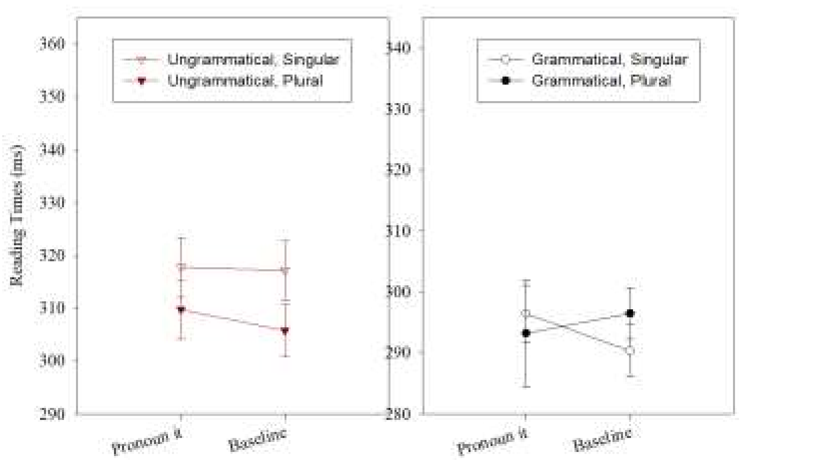


Figure 48. Reading times at the spillover region (*safe*) for all conditions in experiment 9b. Error bars indicate the standard error.

At the critical verb region, a marginal main effect of *Grammaticality* was observed such that ungrammatical sentences were read slowly than grammatical sentences. A main effect of Pronoun *it* was also observed such that baseline conditions were read slower than sentences involving Pronoun *it*.

At the verb spillover region 1, a main effect of *Grammaticality* was observed such that ungrammatical items were read significantly slower than the grammatical counterparts. An interaction between *Local noun* and *Grammaticality* was observed. Further subset analysis in the baseline condition revealed a main effect of *Grammaticality* ($\beta = 18.51, SE=4.42, t=4.19, p<0.001$) as well as an interaction between *Local noun* and *Grammaticality* ($\beta = -16.74, SE=7.33, t=-2.28, p<0.05$), but only a main effect of *Grammaticality* ($\beta = 18.43, SE=4.22, t=4.37, p<0.001$) in Pronoun *it* conditions. Further subset analysis also revealed a main effect of *Local noun* in ungrammatical conditions ($\beta = -9.58, SE=4.46, t=-2.15, p<0.05$) such that plural local nouns were read significantly faster than the singular local nouns in ungrammatical conditions. No further effects were observed in grammatical conditions. This suggests that both Pronoun *it* and the baseline conditions revealed an agreement attraction effect although the effects were larger for the baseline condition as shown in the subset analysis.

At the verb spillover region 2, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower than grammatical sentences.

3.17. Experiment 9c Pronoun it: An Eye-Tracking while Reading Experiment

3.17.1. Participants, Materials and Design

In this experiment, 65 native speakers of English from the Northwestern University community with no history of language/reading disorders participated. Before the experiment, an informed consent form was signed, and the participants received credit (1 credit/45 minutes) for an introductory Linguistics class offered at Northwestern University.

Same critical items were used as in Experiment 9b. Items were distributed in a pseudo-randomized manner so that the experimental items of from the same experiment did not appear adjacent to each other. The experimental items were mixed with 100 filler sentences of similar length and complexity.

3.17.2. Procedure

Similar procedure was employed as in 6c.

3.17.3. Analysis

Similar analysis was employed as in 6c.

3.17.4. Results

<u>Verb Region</u>

In the Total Fixation Time, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower than the grammatical sentences.

Spillover region 1

In the Regression Path Duration, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower than grammatical sentences. A main effect of the *Pronoun it* was observed such that sentences with the Pronoun *it* were read significantly slower than the baseline conditions. A three-way interaction between *Local noun*, *Grammaticality*, and Pronoun *it* was observed. A further subset analysis revealed a main effect of *Local noun* (β = -0.13, SE= 0.05, t= -3.62, $p<0.01$) and an interaction between *Local noun* and *Grammaticality* (β = -0.25, SE= 0.11, t= -2.27, $p<0.05$) in the baseline conditions, but only a main effect of *Grammaticality* in the Pronoun *it* conditions (β = 0.13, SE= 0.06, t= 2.10, $p<0.05$). Furthermore, in ungrammatical conditions, a marginal main effect of *Local noun* (β = -0.12, SE= 0.06, t= -1.97), *Pronoun it* (β = -0.19, SE= 0.06, t= -3.17, $p<0.01$) and an interaction between these two were observed (β = -0.29, SE= 0.13, t= -2.28, $p<0.05$). In contrast, only a main effect of Pronoun *it* (β = -0.15, SE= 0.07, t= -2.10, $p<0.05$) was observed in grammatical conditions.

In the Total Fixation Time, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower than grammatical sentences. A main effect of Pronoun *it* was observed such that sentences with Pronoun *it* were read significantly slower than those in the baseline conditions. Furthermore, a marginal three-way interaction was observed, such that only the baseline conditions showed agreement attraction effect.

Spillover region 2

In the First Fixation Duration, a main effect of Pronoun *it* was observed. In the Regression Path Duration, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower those containing the grammatical sentences.

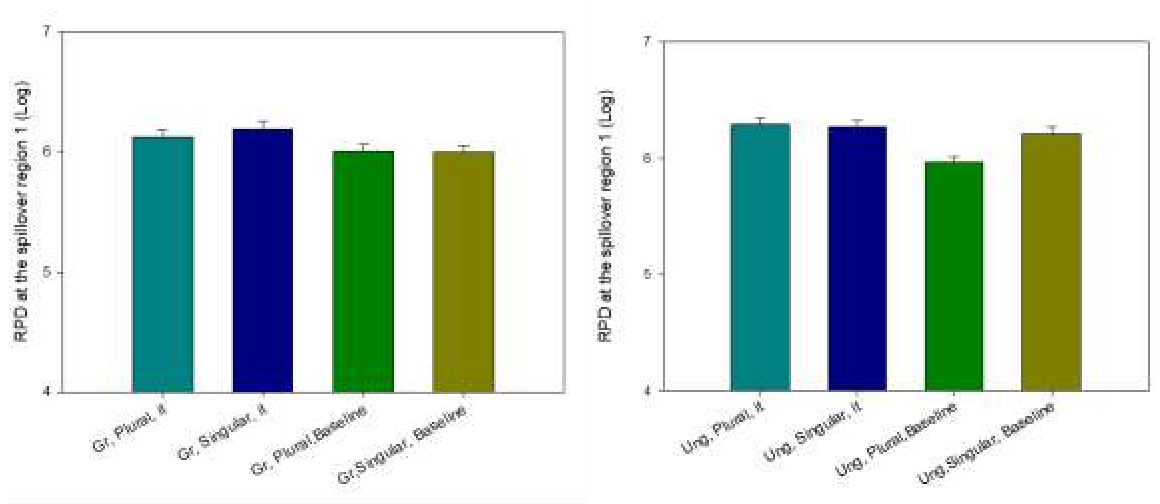


Figure 49. Bar plot for the RPD in grammatical conditions (left) and ungrammatical conditions (right) at the spillover region 1 for Pronoun *it* and the baseline in experiment 9c.

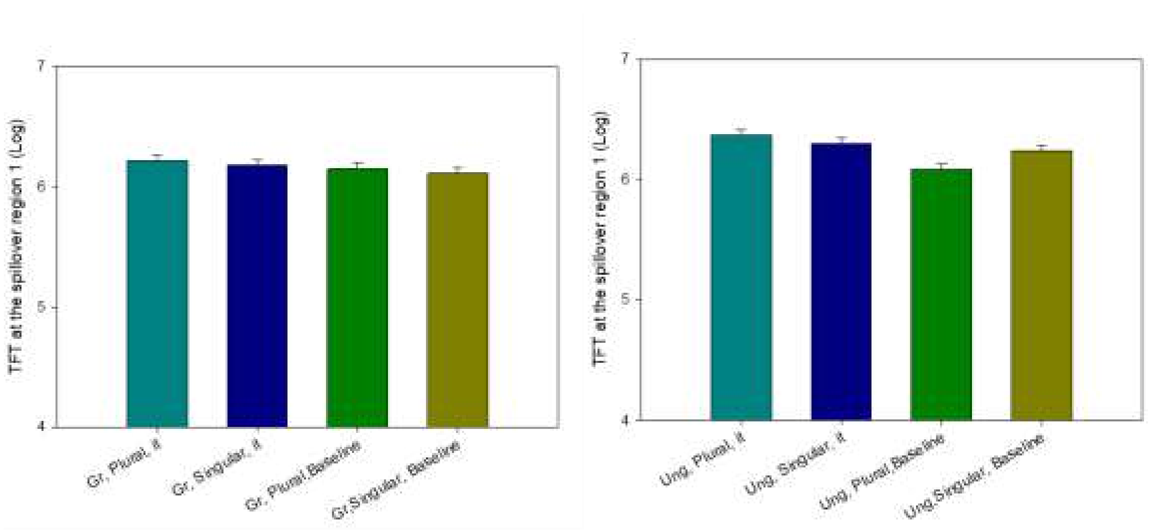


Figure 50. Bar plot for the TFT in grammatical conditions (left) and ungrammatical conditions (right) at the spillover region 1 for Pronoun *it* and the baseline in experiment 9c.

Table 50. The statistical analysis of results for the eye-tracking while reading experiment (experiment 9c) on the verb region, spillover region 1, and spillover region 2.

	Estimate	SE	t	p
Verb Region (*is/are*)				
First Fixation Duration				
(Intercept)	5.47	0.03	181.27	
Local noun	0.02	0.04	0.44	0.67
Grammaticality	-0.03	0.03	-0.90	0.38
Pronoun *it*	-0.04	0.03	-1.13	0.27
Local noun x *Grammaticality*	0.10	0.09	1.06	0.30
Grammaticality x Pronoun *it*	0.10	0.06	1.57	0.12
Local noun x Pronoun *it*	-0.08	0.06	-1.20	0.23
Local noun x *Grammaticality* x Pronoun *it*	0.09	0.13	0.70	0.49
Regression Path Duration				
(Intercept)	5.64	0.04	138.76	
Local noun	0.00	0.05	0.09	0.93
Grammaticality	0.00	0.05	0.09	0.93
Pronoun *it*	-0.08	0.05	-1.46	0.15
Local noun x *Grammaticality*	-0.07	0.09	-0.77	0.44
Grammaticality x Pronoun *it*	0.04	0.09	0.45	0.66
Local noun x Pronoun *it*	-0.10	0.09	-1.15	0.26
Local noun x *Grammaticality* x Pronoun *it*	-0.11	0.17	-0.62	0.54
Total Fixation Time				
(Intercept)	5.63	0.03	195.40	
Local noun	-0.04	0.04	-1.03	0.31
Grammaticality	0.13	0.04	3.23	<0.01 **
Pronoun *it*	0.00	0.04	0.01	0.99
Local noun x *Grammaticality*	0.00	0.08	0.05	0.96
Grammaticality x Pronoun *it*	0.10	0.07	1.36	0.18
Local noun x Pronoun *it*	-0.08	0.08	-1.04	0.30
Local noun x *Grammaticality* x Pronoun *it*	0.12	0.14	0.88	0.38
Spill-over region 1				
First Fixation Duration				
(Intercept)	5.56	0.02	266.60	
Local noun	-0.02	0.02	-0.79	0.43
Grammaticality	0.01	0.02	0.57	0.57
Pronoun *it*	-0.00	0.02	-0.05	0.96
Local noun x *Grammaticality*	-0.02	0.04	-0.57	0.33
Grammaticality x Pronoun *it*	-0.01	0.04	-0.34	0.74
Local noun x Pronoun *it*	-0.070016	0.04	-1.74	0.08
Local noun x *Grammaticality* x Pronoun *it*	0.019654	0.08	0.24	0.81
Regression Path Duration				
(Intercept)	6.12	0.05	119.98	

Local noun	-0.07	0.04	-1.94	0.06
Grammaticality	0.10	0.04	2.77	<0.01 **
Pronoun *it*	-0.17	0.05	-3.41	< 0.01 **
Local noun x *Grammaticality*	-0.10	0.07	-1.39	0.17
Grammaticality x Pronoun *it*	-0.04	0.07	-0.61	0.54
Local noun x Pronoun *it*	-0.11	0.07	-1.59	0.11
Local noun x *Grammaticality* x Pronoun *it*	-0.31	0.14	-2.19	<0.05 *
Total Fixation Time				
(Intercept)	6.19	0.05	130.71	
Local noun	-0.01	0.03	-0.16	0.87
Grammaticality	0.09	0.03	3.00	< 0.01 **
Pronoun *it*	-0.13	0.03	-4.03	<0.001 ***
Local noun x *Grammaticality*	-0.08	0.07	-1.44	0.15
Grammaticality x Pronoun *it*	-0.10	0.07	-1.84	0.09
Local noun x Pronoun *it*	-1.11	0.57	-1.70	0.07
Local noun x *Grammaticality* x Pronoun *it*	-0.20	0.11	-1.80	0.07
Spill-over region 2				
First Fixation Duration				
(Intercept)	5.50	0.03	216.81	
Local noun	0.02	0.02	0.99	0.33
Grammaticality	-0.01	0.02	-0.59	0.56
Pronoun *it*	0.06	0.02	2.47	<0.05 *
Local noun x *Grammaticality*	0.03	0.05	0.62	0.54
Grammaticality x Pronoun *it*	-0.01	0.05	-0.27	0.79
Local noun x Pronoun *it*	0.03	0.05	0.64	0.52
Local noun x *Grammaticality* x Pronoun *it*	0.17	0.09	1.85	0.06
Regression Path Duration				
(Intercept)	7.15	0.08	89.62	
Local noun	0.05	0.05	1.02	0.31
Grammaticality	0.14	0.05	2.94	< 0.01 **
Pronoun *it*	0.08	0.05	1.55	0.13
Local noun x *Grammaticality*	-0.11	0.09	-1.21	0.23
Grammaticality x Pronoun *it*	-0.01	0.09	-0.07	0.95
Local noun x Pronoun *it*	-0.03	0.09	-0.36	0.72
Local noun x *Grammaticality* x Pronoun *it*	0.20	0.19	1.06	0.29
Total Fixation Time				
(Intercept)	6.09	0.06	96.90	
Local noun	0.00	0.03	0.14	0.89
Grammaticality	0.01	0.03	0.37	0.72
Pronoun *it*	0.02	0.03	0.69	0.50
Local noun x *Grammaticality*	-0.01	0.06	-0.16	0.88
Grammaticality x Pronoun *it*	0.07	0.06	1.13	0.26
Local noun x Pronoun *it*	-0.09	0.06	-1.49	0.14
Local noun x *Grammaticality* x Pronoun *it*	-0.05	0.12	-0.37	0.71

3.17.5. Discussion of Experiment 9: Pronoun it

Both in the self-paced reading experiment and an eye-tracking while reading experiment, a main effect of *Grammaticality* was observed at an earlier stage at the verb region, consistent with Anaphoric *one*. For the self-paced reading experiment, an interaction between *Local noun* and *Grammaticality* was observed, suggesting that both Pronoun *it* and the baseline somewhat show agreement attraction. However, the results of an eye-tracking experiment showed a lack of agreement attraction altogether. At the spillover region 1, in the RPD, a three-way interaction between *Local noun*, *Grammaticality*, and Pronoun *it* was observed, suggesting that sentences with Pronoun *it* were read differently from the baseline conditions in ungrammatical conditions. The sentences with ungrammatical plural local nouns were read significantly faster than those with ungrammatical singular local nouns in the baseline conditions but not with the sentences involving Pronoun *it*. This further suggests that the ways that the antecedent is represented for Pronoun *it* and the ways it is represented for NPE may be different. Pronoun *it*, similar to Anaphoric *one*, provides a clear singular morphological cue (head noun and the noun category) that matches/overlaps with the features associated with the head noun in the first conjunct. Therefore, the parser should access the head noun when it encounters the pronoun *it*, and will be surprised to observe the ungrammatical verb, *are*. The second possibility is the sequence, *it are,* is very unnatural, and the parser may just give up parsing, leading to no agreement attraction effects for the Pronoun *it* condition.

Pronoun *it* is known to have no structure involved, meaning that the antecedent retrieval is guided by the non-structural cues. Therefore, the parser would primarily be focusing on the head noun, without checking the features of the modifier (local noun) to fix the number mismatch problem.

However, according to some of the syntactic analyses, pronouns are analyzed as determiners, as pronouns are interpreted as definite descriptions and pronouns can co-occur with NPs (e.g., *we linguists*, *you smokers* etc) (Elbourne, 2001, 2008; Postal, 1969). Elbourne (2008) posits the definite NP, *the*, behaves like a licensor where the NP complement is elided which triggers NP Ellipsis. These theories assume DP including pronouns involve deletion. We observed a small interaction between *Local noun* and *Grammaticality* in the self-paced reading experiment for Pronoun *it*. Our results of the self-paced reading experiment are, thus, somewhat compatible with the idea that Pronoun *it* involves definite NPs with ellipsis (Elbourne, 2001, 2008; Postal, 1969).

There are two possible scenarios. First, the source of the plural noun should be coming from the verb. Thus, when the parser encounters an anaphoric element, *it*, the parser would initially access the head information. When the mismatch occurs in terms of the number feature between *it* and ungrammatical verb, *are*, the parser should access the local noun to fix the mismatch problem. The second possibility is that the source of the plural noun could be coming from the plural local noun in the elided NP. If we assume that pronouns involve deletion (Elbourne, 2001, 2008; Postal, 1969), the parser may access the plural noun inside the NP, but this whole NP is deleted as a whole after the agreement calculation, still giving rise to agreement attraction.

However, in the eye-tracking experiment, we did not see any agreement attraction effect. The lack of agreement attraction for Pronoun *it* in an eye-tracking while reading experiment strongly suggests that the Pronoun *it* behaves similar to Anaphoric *one* in providing certain cues (anaphoric NP and singular N). Based on the results from eye-tracking experiment, we can conclude that what is retrieved by pronoun and NPE is different, and that the antecedent retrieval process for NPE is guided by some structural information, and that the retrieval of NPE requires accessing the antecedent with the recourse to the structure, but the retrieval for the Pronoun *it* does not.

The discrepancy between the results from the self-paced reading and the eye-tracking results remain. At this point, we cannot draw strong conclusion about what antecedent retrieval process is like for pronoun. However, we would like to temporarily suggest that the total lack of agreement attraction effect in the eye-tracking experiment suggests that the antecedent retrieval process for pronouns and NPE are different, and the result of the self-paced reading experiment is either certain noise in the experiment or explained otherwise.

3.18. 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 twelve 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 6a, 6b, and 6c, 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 6 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 & Clifton, 2001; Frazier et al., 2000; 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 7 was 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 7 do not support this alternative hypothesis; agreement attraction was not observed in reading times or acceptability judgments 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 8 examined constructions involving Anaphoric *one,* which, like NPE, needs to access and reactivate the antecedent.

Finally, we further tested whether the parser reactivates information about the antecedent without structural information with regards to the distinction between the head and the modifier. To address this last possibility, Experiments 9 examined constructions involving Pronoun *it*, which, like NPE/Anaphoric *one*, engages in accessing as well as reactivating the antecedent. However, like Anaphoric *one*, Pronoun *it* also provides peculiar features of the singular property.

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 recognize 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 recognized, the parser carries out the following processes: (i) the parser retrieves the information associated with the head of the antecedent NP, (*[[$_{head\text{-}N}$ 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 recognizing 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 6), unlike previous research (Lago et al., 2015; Parker & Phillips, 2017; Tanner et al., 2014) and the No Anaphora, Anaphoric *one*, and Pronoun *it* experiments (Experiments 7, 8, & 9), 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.

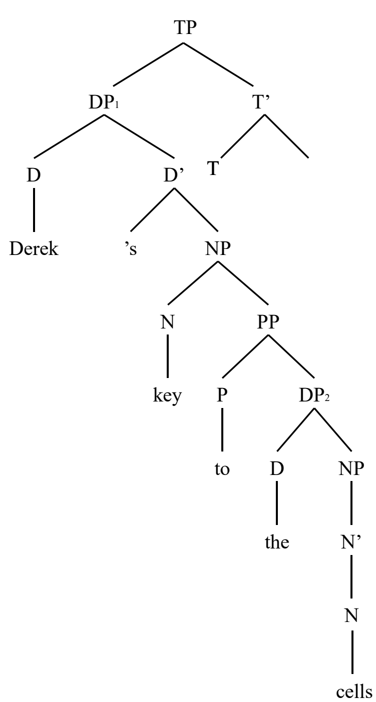
We suggest this difference in time profiles might be attributed to the availability of morphological cues. For NPE, the parser can recognize the ellipsis site when the parser encounters the possessive marked noun and an adverb. Spelling out a possible time course of the recognition of the NPE-site, it 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 recognizes that the anticipation is failed, and also recognize 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 recognizes 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, 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.

We would like to discuss this in the context of the retrieval mechanism and the copying mechanism. First, NPE, Anaphoric *one* and Pronoun *it* should all reveal similar agreement attraction effects. They are all anaphoric elements in which their meaning is controlled by their antecedent. However, we observed an illusion of grammaticality in the NPE experiment. To capture this effect, the parser needs to retrieve/reactivate the structural information. There are multiple DPs existing in the already processed part of the sentence, i.e., the DP that is immediately dominated by TP, and DP located within the PP that serves as a modifier of NP, yet the parser seems to find a DP that contains the antecedent of the NPE-site, namely the DP that is immediately dominated by TP. This suggests that the parser should be able to recognize the structural difference between these two DPs. In terms of retrieval cues, this distinction can be represented as structural features (Lewis & Vashishth, 2005): they can be represented as [+Spec TP] for DP_1 and [+ Sister of P] for DP_2. Specifically, Lewis & Vasishth (2005) assume the maximal projection (XP) as chunks. Each chunk possesses features specific to the positions in the X-bar structure (specifier, head, complement) in addition to other grammatical features such as case/grammatical role (Lewis & Vasishth, 2005: 385).

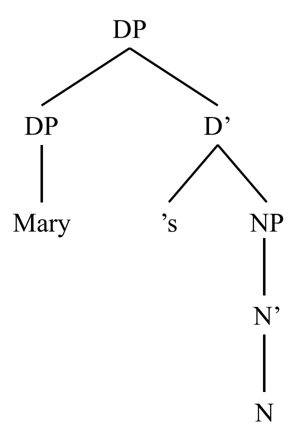
(76)



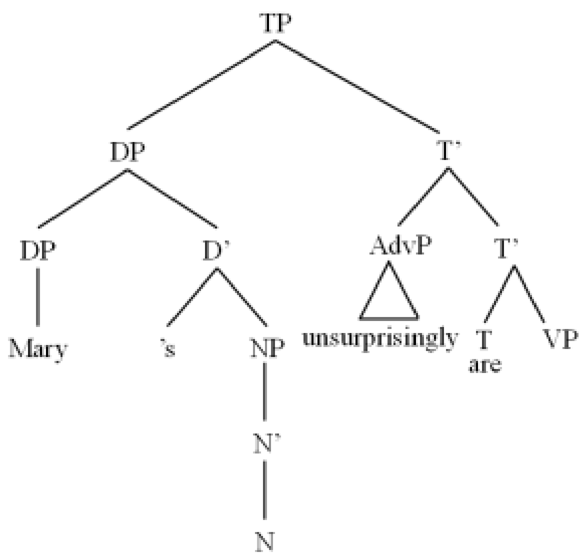
Our results can be captured by a retrieval theory that incorporates the structural information as retrieval cues, i.e., the antecedent retrieval of the NPE-site involves retrieval of structural and morphological information. The retrieval of these different anaphoric elements is guided by structural information which includes the [+Spec TP], head/modifier information, and category information (Lewis & Vasishth, 2005; Van Dyke & Lewis 2003; Van Dyke & McElree, 2011) in addition to the number information.

Let us now look at the time course of the processing of NPE constructions. The NPE-site retrieves the syntactic structure of the antecedent, not just the head noun. In other words, the whole NP serves as a retrieval cue. When the parser encounters *Mary's*, upon the top-down prediction of the NP-node, and an adverb, *unsurprisingly*, the parser builds the VP structure. This is due to the grammatical incompatibility between the NP and an adverb, which leads to the structure of sentence which involves NP and VP (*[$_{S}$ [$_{NP}$ Mary's [$_{NP}$]] [$_{VP}$ [$_{Adv}$ unsurprisingly] [$_{VP}$ ...]]]*) as illustrated in Step 2.

(77) Step 1.



(78) Step 2.

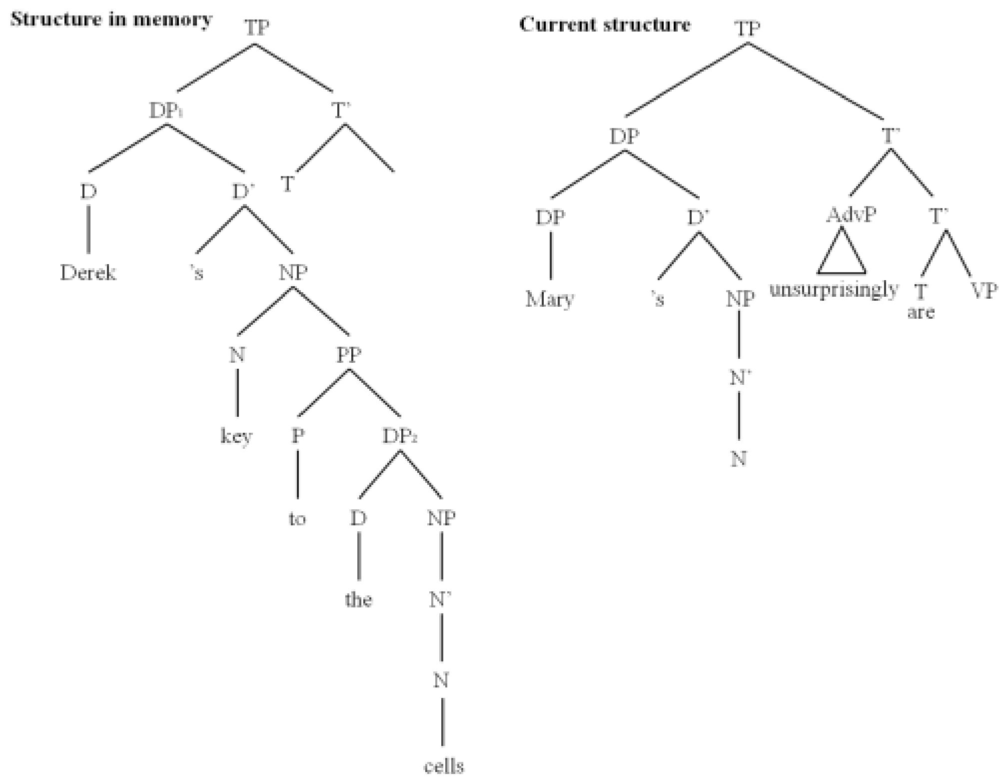


Upon encountering the possessive marked noun and an adverb, the parser recognizes the ellipsis site, due to the presence of an adverb and the lack of head-noun[40]. Upon recognizing the NPE-site, the parser retrieves the whole NP. But the parser first needs to inspect a wider structural environment such as Spec_TP and its structural and relational information. This would lead the parser to recognize that the DP_1 is embedded in the Spec_TP node.

(79) Step 3.

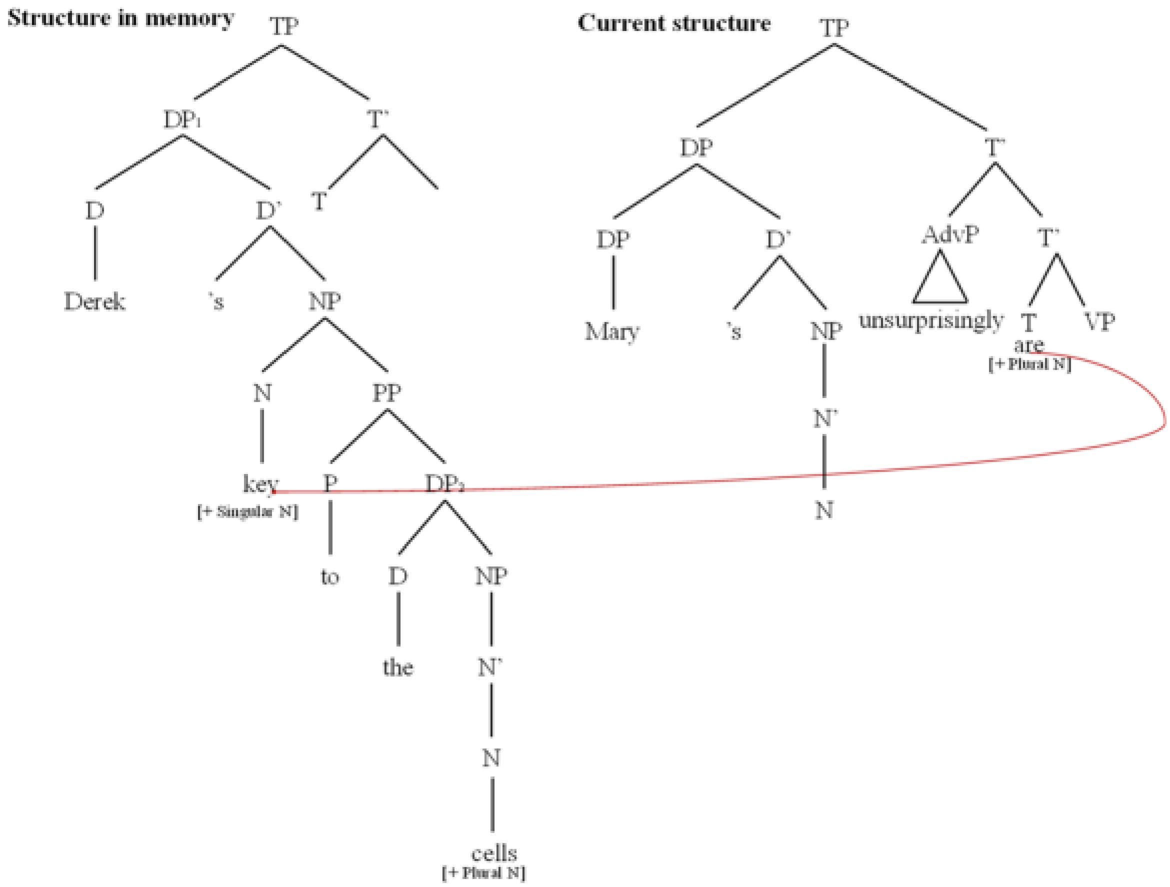
---

[40] It is true that an adverb can be within the DP serving as a modifier of the AP (e.g., *Mary's unsurprisingly small keys*). However, we assume that an adverb as a VP modifier is more likely to occur compared to an adverb as the AP modifier.



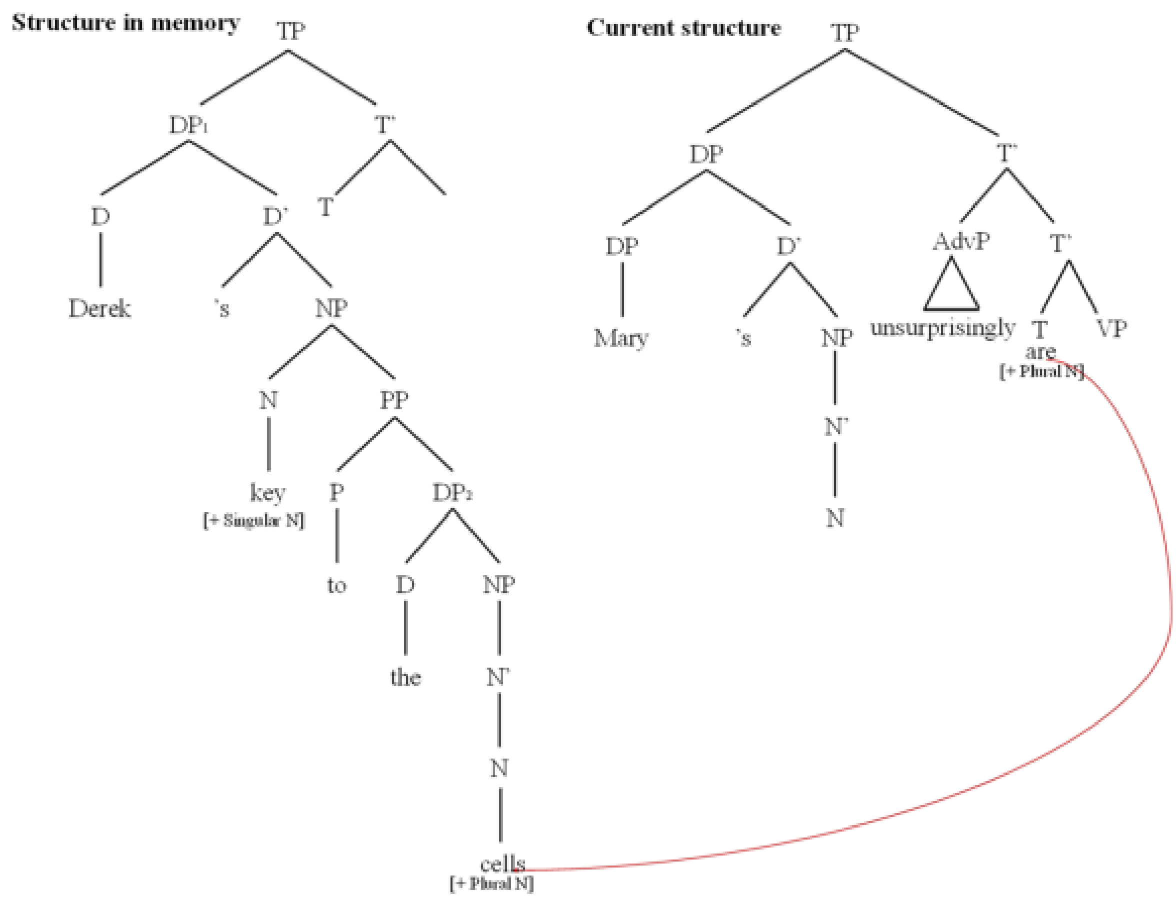
The whole DP inside the DP_1 entails the NP with the PP inside. Based on this information, the parser first computes the agreement relation between the head noun (*key*) and the verb (*are*), i.e., computes the agreement between the head noun and the verb.

(80) Step 4.



The parser would look for the agreeing noun as a second attempt (*cells*) and the plural verb and the plural local noun match with regards to the number feature.

(81)



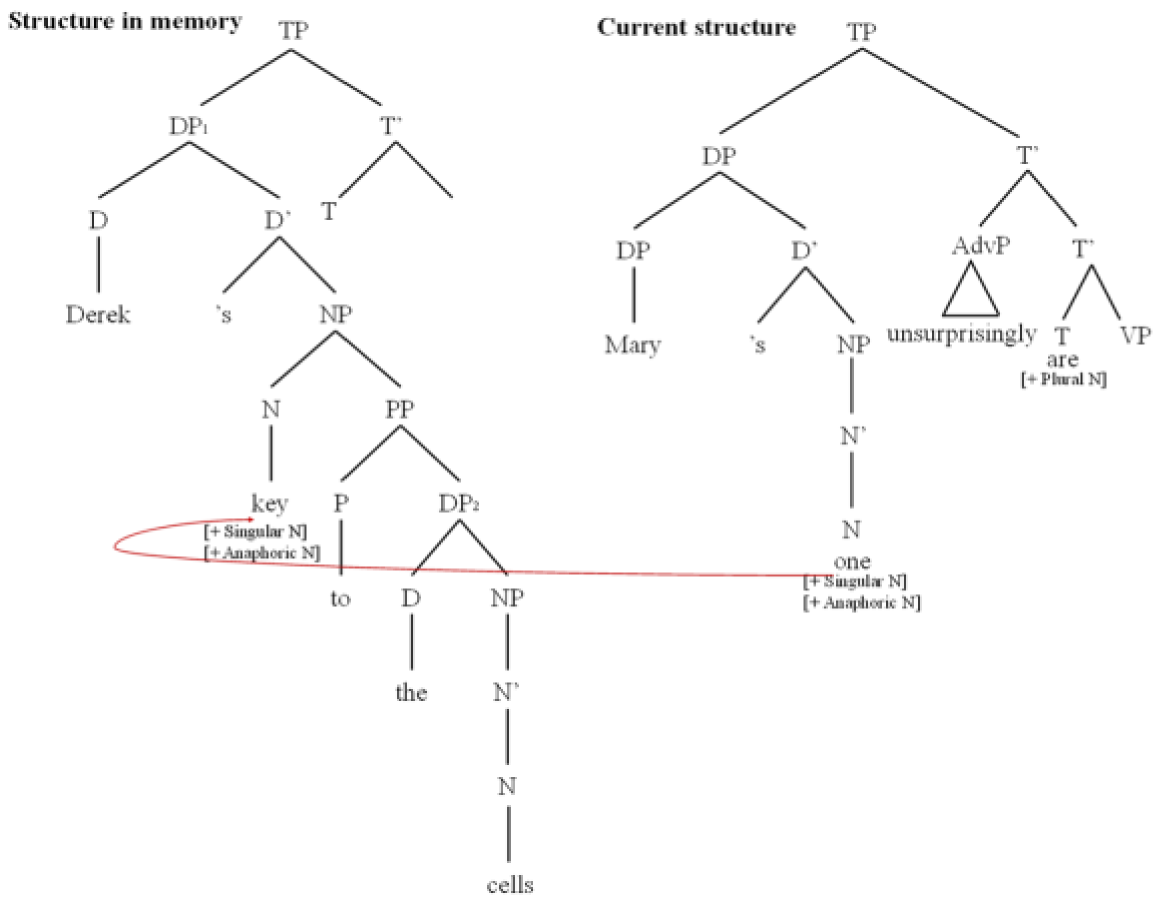
On top of the process involved in agreement calculation, resolving ellipsis requires the recovery process. This extra step of processing involved in the recovery of ellipsis can result in a grammaticality effect at a later stage than the verb for the NPE-site. Thus, from the retrieval perspective, a cue-based retrieval mechanism needs to posit the structural information to be encoded as retrieval cues. This may include the [+Spec TP] and the structural distinction.

Let us examine Anaphoric *one* and Pronoun *it* from the retrieval view. From the retrieval point of view, we do not expect the differences between NPE and Anaphoric *one*/ Pronoun *it* as all three anaphoric elements need antecedents to achieve the meaning. If the antecedent is retrieved in the same way in these three constructions, it is not clear why we do not observe an agreement attraction effect for all these elements. The parser could retrieve the local noun when the head noun and the verb mismatch in number feature if the whole antecedent is retrieved. In the cases of Anaphoric *one* and Pronoun *it*, when *one/it* is recognized, *one/it* can trigger the retrieval immediately and calculate the agreement with the verb. Therefore, similar to NPE, the parser would retrieve the DP_2 only when the number features of the DP_1 and the verb do not match. In either case, the retrieval story could predict an agreement attraction for Anaphoric *one* and Pronoun *it*. Then why do we not observe an agreement attraction for *one* and *it*? There are three possibilities. First, these elements provide a strong morphological cue that marks an anaphoric element with the singular N, noun category, and the head noun. Therefore, the parser would not retrieve the local noun even if the head noun and the verb mismatch in terms of the number. But in this scenario, we are unsure why the retrieval cues provided by the verb play no role.

The second possibility is that the sequence of *one are* or *it are* are unnatural sequences and the parser would just give up parsing. In this case, Anaphoric *one* simply does not retrieve the antecedent but rather is surprised by the mismatching verb. Therefore, the parser may not compute the agreement relation, suggesting that both an ungrammatical plural local noun and a singular local noun would yield similar processing profiles.

Third, it is the case that the head noun in the antecedent NP is more prominent because it is the head of the phrase. Thus, the parser reactivates primarily the information of the head.

(82)



In fact, the ACT-R model which involves the syntactic structure building would predict that the head noun has higher activation values than the modifier as it is more prominent in the hierarchical structure (Lewis & Vasishth, 2005). If this is the case, we would observe the same effect in NPE, Anaphoric *one* and Pronoun *it* (no illusion of grammaticality) because the head noun is very prominent. At the very least, retrieval cues used to retrieve these different anaphoric elements need to involve the representation with regards to the head noun and the modifier.

Finally, what if Anaphoric *one* and Pronoun *it* can only use semantic features as retrieval cues as they are deep anaphora? In this case, the search space of the antecedent of these elements should be restricted by [+Spec DP]. However, these elements are primarily cuing for meaning features of the antecedent, not the structural features because they are pronominals. Because these pronominals are deep anaphora which are semantically and pragmatically controlled (Hankamer and Sag 1976), it is plausible that these pronominal elements primarily refer to and reactivate the meaning (semantic or discursive) information of the antecedent. In this case, the deep vs. surface anaphora kind of differences should be encoded as retrieval cues of these anaphoric elements.

The copying theory (Frazier & Clifton, 2001) would copy the DP in the antecedent, which already signals the [+Spec TP], i.e., the parser copies the structure of the antecedent DP into the NPE-site and builds the structure of NP within the NPE-site. If the parser does not build the structure at the ellipsis site, the parser may not be able to make subtle distinctions between different DPs involved in the antecedent (DP with the Spec TP and DP which is the complement of PP). Therefore, the copying mechanism should be sensitive to the type of anaphoric elements. In case of ellipsis, the structure must be recovered, but if it involves deep anaphora, only the meaning representation is necessary and hence no copying is triggered. One thing to note is that the copying theory cannot predict why we do not observe a similarity-based interference effect. If the whole DP is retrieved for the NPE and the Pronoun *it*, and pro-N-bar for Anaphoric *one*, the parser would sometimes retrieve the embedded NP in the DP_2 as there are two different DPs. However, pure copy theory would predict that when there is no structure of the antecedent, no copying should be involved.

In this sense, copying makes quite straightforward predictions in accounting for the differences between NPE vs. Anaphoric *one*/Pronoun *it*. If copying plays a role, Pronoun *it* and Anaphoric *one* should also retrieve whole content of the antecedent NP. For the

Pronoun *it*, this would involve the DP involving pronouns and for Anaphoric *one*, this would involve pro N-bar. Yet, we are unsure why we do not observe an agreement attraction for Anaphoric *one* and Pronoun *it*. We show that both Anaphoric *one* and Pronoun *it* do not involve structure, as they are deep anaphora. This means that no copying is involved for these anaphoric elements. However, they provide a strong morphological cue (singular NP, category information) that overlaps with the features of the head noun. This would lead to the failure of the retrieval of the embedded NP inside the DP. Even in this case, we need to assume that the peculiar feature of the singular property of Anaphoric *one* and Pronoun *it* plays an important role. Therefore, the copying approach would predict sharp differences in ellipsis versus other types of anaphoric elements. In terms of the retrieval mechanism, we need to assume some kind of structural information in content-addressable memory, which would require quite sophisticated retrieval mechanisms that would refer to the structural representation (e.g., the Spec TP feature). Furthermore, such a retrieval theory would have to assume there are two different DPs and that the DP to which the parser needs to initially attend is not the DP that is the sister of P. Also, morphological cues as feature components ([+head] feature) need to be incorporated in order to account for the differences between NPE vs. Anaphoric *one*/Pronoun *it*.

4. Conclusion

4.1. General Conclusion

This dissertation addresses the real-time processing of ellipsis constructions (backward dependency formation) and the online WhFGD formation (forward dependency formation), to reveal the mechanism working behind the dependency formation process, and how this process interacts with various components in memory representations.

First, we examined whether different wh-phrases are maintained differently, and at which point they are released from memory. We found that different wh-phrases are released from memory once they are linked to their controlling element. Second, under the assumption that the retrieved information can inform us of what information is accessed, we showed that different filler types and configuration types give rise to different degree of elicitation in agreement attraction: readers access and retrieve rich kind of information such as the representation of the head and the modifier as well as the category information. We showed different reactivation profiles with regards to information that is released from memory and later put into maintenance again versus those that are maintained over the course of the dependency formation process.

We also investigated the antecedent retrieval mechanisms of elliptical constructions (NP Ellipsis) in comparison to other anaphoric elements (Anaphoric *one*, Pronoun *it*, and No Anaphora conditions). The resolution of NPE involves recognizing the ellipsis site, and accessing and retrieving the content of the antecedent to the NPE-site. This is characterized as the retroactive backward dependency formation. We attempted to understand what kind of information associated with the antecedent is retrieved at the ellipsis site, by utilizing an agreement attraction effect as a probe. What we found was the retrieved information differed based on different anaphoric elements being processed, and that the recovery of NPE-site involves structural information (the distinction between the head and the modifier) whereas the recovery of other anaphoric elements (Anaphoric *one*, Pronoun *it*, No Anaphora) does not. Based on these findings, we concluded that the antecedent retrieval process of NPE is governed by structural information, unlike that of other anaphoric elements.

These studies contribute to the theory of sentence processing, memory, and parsing in general. In terms of the maintenance, the theory of sentence processing should assume the maintenance component which assumes a special memory state in addition to the retrieval component in our current memory architecture: there must be a stack like short-term storage device (memory mechanisms such as hold cells or stacks) which is used for maintenance (Frazier, 1987; Wanner & Maratos, 1978). The important point is that in the maintenance component, we have to posit a stack-like special memory state which can hold not only words or categories, but also structured object. In this sense, the maintenance is understood as something like stack, and the parser can access the element and put the element into the stack in terms of the structure building process. Furthermore, on top of the stack-like mechanism, we have to posit some longer-term storage where processed materials are stored and processed things are retrieved from (Lewis & Vasishth, 2005; McElree, 2006). From this perspective, we could understand the maintenance as the state where structural information is put in a stack-like special memory device, until it forms the dependencies and reduces structure from the stack. What happens when the parser encounters the tail/licensor of the dependency? The release would indicate the state when the dependency is formed by means of encountering the tail/licensor of the dependency, and thus the output of the parsing goes into the storage. In other words, when the structure is built and the dependency is formed, these structures/elements in the dependency are reduced from the stack and put into the different memory state (storage). At the later point of the sentence, if there are some elements that provide retrieval cues, then the retrieval operation is triggered and hence the parser searches for the elements that have features that match the retrieval cues, and such element is put back into the maintenance (something like stack) again. Therefore, retrieval should refer to the processes of accessing and reactivating the processed material in the storage component and putting back to the stack for parsing.

We conclude that maintenance, release, and retrieval all play crucial roles in the course of the resolution of WhFGD, and ellipsis constructions, and these components dynamically interact with structure building.

4.2. Specific Conclusion

This dissertation addressed the question of how various components in our memory representation interact with the syntactic structures. We showed that maintenance, release and retrieval should be taken into consideration in order to capture the mechanisms working behind the resolution of different dependencies.

First, we addressed the question of whether the maintenance (storage) component exists by examining the processing of different kinds of wh-elements (*who* in the subject position, *who* in the object position, *how* and *why*). Some of the syntactic studies suggest that these wh-phrases are licensed and thus linked to different structural positions, where *who* in the subject position and *why* are linked to the TP, whereas *how* and *who* are linked to the VP and V, respectively. We employed storage cost as the way to understand how different wh-phrases interact with the maintenance. We presented evidence that *why* and *who* in the subject position, unlike *how* and *who* in the object position, are released from memory as soon as they reach the controlling element. From the maintenance perspective, the parser attempts to release the wh-phrase from memory once wh-phrase is licensed by the grammatical requirement of different kinds of fillers. Different processing profiles of wh-phrases with respect to different dependency lengths suggest that the parser stores different wh-phrases in memory and releases once the licensor is recognized.

To account for this data, we are assuming a memory architecture where the wh-phrase is put into a special state as maintenance in some device like stack in addition to the storage component. Thus, some kind of device is where maintained element is put, and the storage is where elements to be retrieved are stored. We showed that the structural object is actively maintained in terms of structure building process in the stack-like special memory state as maintenance. When the parser completes the dependency, by linking the wh-phrase with the gap, the parser releases this wh-phrase from the stack. In this sense, forgetting should imply reducing something from the stack and putting into in the storage. If the parser needs to retrieve the already processed element, the parser needs to retrieve and access information in the storage and put back into the stack to undergo structure building process.

Then we could ask whether the differences in the storage costs for different wh-fillers can be explained by retrieval theories without positing the maintenance component. Yet, if we assume that only retrieval plays a role, we need to assume a large number of adhoc retrieval cues. For example, TP should encode *why* as retrieval cues, where *why* should be dependent on TP, but TP should not be dependent on *why*. Similarly, VP should encode *how* as retrieval cues, where *how* is dependent on VP, but VP is not dependent on *how*. Finally, V should encode *who* as retrieval cues, where *who* is dependent on V, but V is not dependent on *who*. On the other hand, if we assume that *why* should be dependent on TP and should be licensed by TP (where *why* is the predictor of TP) and so on for other wh-phrases, our results follow naturally. In this sense, the maintenance component can capture the storage costs involved in the processing of different wh-phrases.

We also want to note that *why* and *how* should be processed differently in parsing to account for our data. By selectional/subcategorization feature and X' theory we are assuming, if TP is built then VP is entailed. If this is the case, *how* can be connected to the predicted VP. This means that *why* and *how* which are licensed by TP and VP respectively should not show differences in terms of the processing, because once TP is built, *how* can be released from maintenance. However, we showed that at the most deeply embedded NP in a relative clause, *why* was read significantly faster than *how*, as *why* did not contribute to the additional processing cost at this relevant point. To account for the data, we argued that given the left to right processing, there should be a parsing step where *why* and TP exist, but VP does not exist. In other words, there should be a state where *why* and TP are in the short-term storage, and this stack does not include VP. This can potentially be achieved by the combination of some context free grammar and the arc standard left corner parser which allows partial structures to float around without integrating into the existing or bigger structure. These results strongly suggest that the theory of working memory and sentence processing should assume the maintenance component in the online dependency formation.

Based on the findings that the parser indeed maintains some content of information in memory, we addressed the following question: to what extent is the filler maintained and how does the maintenance affect the subsequent retrieval event? We tested different types of configurations as well as different fillers to understand what kinds of information is maintained relatively well and what kind of information is released from memory and hence needs to be accessed again. To understand what information is accessed at the gap position, we observed the strength of an agreement attraction elicited for different types and configurations during WhFGD formation process. We presented evidence that (i) the readers maintain quite fine-grained information associated with the wh-filler such as the representation of the head and the modifier, as well as the category information, and that (ii) the information that is once released from maintenance and reactivated again differs from information that is maintained. Specifically, in the coordinated WhFGD, information is released from memory and then reactivated again upon encountering the coordinative connective *and* by means of grammatical constraints (Wagers & Phillips, 2009). Upon encountering the coordinative connective *and*, the parser forms another dependency actively searching for the upcoming gap in the subsequent conjunct. Since information of the filler is once released from memory, information of the filler is accessed and retrieved at the later stage and in a lesser degree relative to information that is not released.

Therefore, in order to account for the differences observed for active filler vs. reactivated filler, the maintenance component is crucial; once the elements are put into the maintenance device, it should be less susceptible to decay or should not be susceptible to decay. In this sense, different memory dynamics should be employed for the elements in the maintenance device and those in the storage. Following this logic, even if we add the distance between the wh-filler and the licensor, the elements in the maintenance device should not be impacted but the elements already released and put into the storage should be harder to access (Wagers & Phillips, 2014).

Chapter 3 aimed to uncover the retrieval mechanism working behind the resolution of NPE in comparison to other anaphoric elements (Anaphoric *one*, Pronoun *it*, No

Anaphora). As we have discussed in the syntax part, the crucial differences between NPE and these anaphoric elements are that NPE has an internal structure that parallels the antecedent. Thus, during the backward dependency formation, upon encountering the NPE-site, the parser would search for the antecedent to achieve the interpretation. For NPE, when the recovered noun head and the verb mismatch in terms of the number feature, it would look for another noun (local noun). This indicates the parser is sensitive to the structure of the NPE, and hence retrieve structural information at the NPE-site. On the other hand, Anaphoric *one* and Pronoun *it* did not show agreement attraction effect. This is because Anaphoric *one* and Pronoun *it* offer strong morphological cues which indicate the singular NP. Furthermore these pronominal elements are deep anaphora and they are semantically and discursively controlled. Thus they refer to semantic/meaning representation of the antecedent, and reactivate semantic/meaning representation of the antecedent. As a result, the parser would not search for another noun despite the mismatch between the head noun and the verb.

For the NPE, what would happen when an element is released? Would there by any changes in terms of the representation? Even when the element is released from maintenance, and put into the storage component, it is likely that the syntactic structural information remains intact. Thus, the structure built by the parser in stack should be put into the storage as it is, without the changes in the representation. We also want to note that the antecedent should be retrieved as a whole instead of the cases in which the agreement is computed first with the head noun and the local noun is accessed secondarily.

Based on these findings, we argue that maintenance, release and retrieval play crucial roles in the resolution of WhFGD and in ellipsis. How do different ways of forming dependencies relate to various components in memory representations? We discussed earlier that the resolution of WhFGD involves a proactive forward dependency formation, where the dependent element is located before the controlling element. On the other hand, the resolution of ellipsis involves a retroactive backward dependency formation where the parser accesses and recovers the content of the antecedent after the dependent element is processed. Thus, during the resolution of WhFGD, the parser needs to maintain some aspects of information and releases the wh-phrase once it reaches its controlling element. The released information can then be retrieved again. Thus, WhFGD crucially involves maintenance, release, and retrieval over the course of the dependency formation process.

In the processing of elliptical constructions, readers do not anticipate the dependent element that appears later in the sentence. This means that when the reader encounters the ellipsis site, the reader needs to access and retrieve information to the ellipsis site. An important point of our findings in this study is, the information retrieved by the anaphoric elements is different depending on what type of anaphoric elements is processed. In other words, the parser can successfully identify the antecedent by employing grammatical environment as well as grammatical/structural information as retrieval cues to narrow down the antecedenthood.

How do our studies inform us of the theory of memory architecture? We show that there must be a stack like short-term storage device (some mechanisms like stack) used for maintenance (Frazier, 1987; Wanner & Maratos, 1978). This means that in addition to the storage component in which already processed materials are stored and retrieve-to-be materials are stored, there should be a stack like component where the structural information is stored until the dependency is completed. The parser is taking words and categories and putting them into the structure in the maintenance device. If we do not assume this stack like component, and assume that the parser solely accesses the information based on the cue-based feature-matching operation that matches the retrieval cues in memory without building the structure, we cannot account for our results where different WhPs are maintained differently and where different filler types and configuration types contribute to differences in terms of the retrieved information as indexed by an agreement attraction effect. Therefore, our studies show that the maintenance component is working crucially in the resolution of WhFGD and that the theory of retrieval mechanism needs to handle such structural/grammatical information as retrieval cues.

Then, is this maintenance component (stack) employed only to store wh-questions? We argued that one of the arguments for the maintenance is an active dependency formation. Maintaining some elements in memory consumes memory/maintenance resources and thus the reader attempts to release the maintained element from stack as quickly as possible. Previous studies have shown that cataphora processing exhibits an active dependency formation (Kazanina et al., 2007). This means that cataphora processing may also be utilizing the maintenance device, the stack. If this is the case, a cataphoric pronoun should also be put into the maintenance device during the search for the antecedent, and when the antecedent is identified, it would be released from memory. By extension, other forward dependency formations should exhibit active dependency formation, and therefore should also utilize the stack device, similar to the processing of wh-phrases.

Finally, how about the wh-phrases in wh-in-situ languages? Do they make use of the maintenance device in storing wh-phrases? This may differ depending on different languages. For head-final languages, like Korean and Japanese, C which binds the wh-phrase is located to the right of the wh-phrase. When the wh-phrase is encountered, the readers would look for the interrogative C, exhibiting active dependency formation effects

(Miyamoto & Takahashi, 2004). Therefore, we predict that the wh-phrase in these languages should be put into the maintenance device (stack). On the other hand, in the head-initial languages like Chinese, the licensor (the scope indicator such as [+Q, CP]) occurs prior to the wh-phrase (Xiang, Wang, & Cui, 2015). This suggests that the parser would retrieve the licensor (the scope indicator) after encountering the wh-phrase, which occurs after the complementizer. Therefore, in Chinese, the in-situ-wh-phrase should not be put into maintenance.

We have shown that the maintenance component is crucial in addition to the retrieval component in the way that these memory representations interact with syntactic structures. The dependent element in the dependency could have a special marking device that short-term store the elements, but the important point is that the dependent element needs to be stated in a special manner until the licensor is encountered. The differences of the retrieved information depending on the dependency types and the storage cost effects of different wh-fillers follow naturally if we assume both maintenance and retrieval components are actively at play in the online dependency formation. Thus overall, both the maintenance and retrieval process are heavily constrained by grammatical information associated with the elements that engage in dependency formation.

4.3. Future Direction

We have shown that the parser should work in a short-term memory architecture like stack for maintenance in addition to the storage component where the processed elements and retrieved-to-be elements are stored. We need more fully explicit implementation with regards to what the parser should be and how the memory representation/architecture should be like, under the sentence processing mechanism by utilizing computationally explicit models that bear on our underlying cognitive mechanisms.

More specific future directions would be to understand how semantic information can be deployed as retrieval cues (Cunning & Sturt, 2018) and specifically what constitutes semantic cues. One way to test this is to examine whether the parser salvages the mis-parse when contradictory semantic information is encountered. This can be tested by manipulating the adjective (e.g., same or non-contradictory adjectives vs. neutral adjective) as in *John's big kettle's small/red lid was gorgeous but Mary's was big*. In this construction there can be two different potential antecedents for the NPE-site. First, the whole NP can serve as the antecedent, which includes *John's big kettle's small lid* where the head noun is the *lid* rather than *kettle*. On the other hand, given the plausible analysis that *big kettle* form an independent constituent within NP, *big kettle* could in principle serve as an antecedent for the NPE-site as well. Given that *kettle* functions as a modifier and located farther from the ellipsis-site but *lid* is the head of the whole NP which is located near the ellipsis site, it is plausible that there can be a preference for the parser to initially match up the NPE-site with the closer NP *small lid*. One could examine whether the parser primarily resolves NPE with the closest NP. When the retrieved NP involves an adjective that contradicts the adjective in the main clause (e.g., *Mary's small lid was big*) the parser cannot interpret the ellipsis site. In such a case, it is possible that the parser reanalyzes the ellipsis site and changes the antecedent. If evidence of reanalysis is observed, then it means that the parser has a bias to resolve NPE with the closest potential antecedent, and that the parser retrieves modifiers like adjectives during NPE resolution. That is, if the meaning of an adjective is retrieved in such a way, then we could think about whether current retrieval theories of online ellipsis resolution predict this pattern, and if not, what modification is necessary in the current retrieval mechanisms.

As an extension of how different types and qualities of WhFGDs contribute to different retrieval profiles, I aim to investigate whether gender information can be retrieved in Parasitic Gaps (PG) by using maintenance and retrieval as a probe. It has long been known that the gap inside the adjunct or subject islands can be licit when there is another gap positioned in the main clause, which is called parasitic gap (Engdahl, 1983). PG constructions (e.g., *Which boy did you criticize t before making himself a cup of coffee?*) provide an excellent testing ground here because the parser releases the wh-filler and reactivates the wh-filler only after the second gap is processed. Furthermore as shown in (83), illicit gap (PG) can precede the gap that licenses PG as in (83) and thus the parser does not expect a PG to be present.

(83) Which books about himself did John file t before Mary read?

Given that the PG is optional, the parser may not maintain the wh-element and reactivate the wh-element at the verb. If we assume that wh-phrase is released from memory, only category information can be retrieved and not the content of the wh-phrase. If this is the case, we expect information such as gender to be decayed and lost, making it hard to retrieve such information. Therefore, the reactivation at the second gap would lead to very coarse information where gender information is not being retrieved. This will be revealed by the lack of gender mismatch effect due to the retrieval of decayed information. I suggest comparing adjunct PG directly to subject PG which is similar to the adjunct PG in the sense that the gap is optional and hence the parsers' postulation of the gap is not guaranteed. However, the parser may actively insert the gap in subject island when grammatically sanctioned, i.e., the gap is licensed as a PG (Phillips, 2006).

(84) Which picture of himself did every girl who saw say Jasmine loved t?

Subject PG precedes the gap in the main clause which is the gap that the wh-phrase is linked up to. If this is the case, linking the wh-phrase to the subject PG does not establish a wh-gap dependency and thus the wh-phrase should be maintained when the subject PG is processed. If this is the case, then information retrieved at the subject PG should be finer-grained because the wh-phrase is maintained. Therefore, we expect a stronger gender mismatch effect, as gender information is likely to be maintained well. If the differences in terms of the active maintenance of the information of the gender match/mismatch is observed, this would in turn suggest that the parser inserts the gap actively retrieving information of the filler at the first gap. Following this reasoning I would like to compare the retrieval of adjunct PG with subject PG and use the coordination construction (and relative clause) as a baseline so that all these constructions differ with respect to information being maintained and hence affect the retrieved information.

I would also like to extend my current research to a better understanding of the nature of retrieval of the antecedent of the gaps left by the movement of wh-elements vs. the antecedent of the ellipsis-site. This will provide us a better understanding of the nature of retrieval mechanism and whether it is crucially constrained by grammatical information or not. Both gaps and ellipsis should show similar retrieval profiles as they are both anaphoric empty elements in which their interpretation is dependent on their antecedent. However, despite their similarity, there are clear grammatical differences between these two (Fiengo & May 1994; Hankamer & Sag, 1976). A pronoun embedded within the ellipsis site can pick up different antecedent from the pronoun in the antecedent clause (e.g., *John loves his cat and Jasmine does love her cat too*), which is called sloppy identity. It has been shown that sloppy identity is tolerated in ellipsis, but not in the gap. This difference can be attributed to the different structures associated with the ellipsis site and the gap. The gap is the exact copy of the moved element, but the ellipsis is not the exact copy of the antecedent. In other words, how the ellipsis site and the gap are related to the antecedent is grammatically different. This particular property presents challenges to the theory of cue-based models as they are usually not compatible with constraints that rely on structural configurations. Cue-based models may thus have difficulty implementing the differences between gaps and ellipsis in terms of sloppy identity. This investigation could deepen our understanding of whether the parsers' retrieval system is either syntactically well-formed or its operation is simply governed by a cue-matching mechanism. These types of studies allow us to understand what kind of retrieval cues are utilized and what kind of retrieval cues are not utilized, and how certain kinds of grammatical properties (e.g., difference between the ellipsis site and the gap) are implemented in the retrieval mechanism.

Cross-linguistic investigation that integrates cognitive approaches could provide us with a better understanding in maintenance and retrieval components in general. Different languages encode different morphological and semantic features in language expressions such as words and phrases. Looking at typologically different languages can inform us of (i) how rich representations and constraints of diverse languages are deployed in sentence processing and (ii) what sort of features are deployed as retrieval cues. In this respect, if we find differences between English and other languages, we can potentially explore which features are employed as the retrieval cues and which features are not. I propose investigating Korean as a way to examine the underlying mechanisms of encoding, maintenance, and retrieval processes, with a focus on special properties found in that language useful for this task: honorific markers in subject-verb agreement and case information on nouns. Honorific and case information are excellent candidates because they employ different kinds of morphosyntactic properties which can signal the presence of certain dependencies, serving as a cue for online dependency formation. Studies in English show that the parser mostly utilizes grammatical features expressed in word forms (like person, number, or gender), but it is not clear whether more abstract semantic features can be used as cues for different types of dependency. For example, Wagers & Phillips (2014) suggest that the syntactic category information of the filler is maintained in memory, while its lexical or semantic content decays to be retrieved later. But that proposal is underspecified as to what counts as *semantic* versus *syntactic* content, and recent research (Chow & Zhou, 2019) has challenged their evidence in English. In this respect, a comparison of the Korean honorific and case system can be a good test of the memory retrieval mechanism. Specifically, I am interested in what kind of structure the parser would prefer to build in Korean when encountered with subject-verb honorific mismatch, as in (85).

(85) [S [NP Subject_non-honorable] [S [NP null-subject] [VP V-honorific ...]] [VP V]]

In Korean, verbal morphology is sensitive to the socio-cultural status of the subject NP. Thus, when the subject is an honorable person, the verb bears corresponding honorification morphology. However, it is sometimes possible for an un-honorable subject to be followed by a verb with an honorific marker. In such cases, however, there must be an implicit non-overt subject in between the overt subject and the verb, and the verb's morphology is controlled by this implicit subject. There are multiple possible structures that involve an implicit subject: one is to have a relative clause structure, and another is a simple embedded clause with a null pronoun subject. This property leads to the following question: When the subject-verb honorification mismatch is recognized, and thus an implicit subject is to be inserted, what structure would Korean readers build (simple embedded clause with an empty subject or a complex noun phrase structure involving a relative clause)? It is known that when the parser recognizes an anaphoric element (like a pronoun or a gap) the parser tries to find its antecedent within the same sentence where an anaphoric element is recognized (Aoshima et al., 2004; Kazanina et al., 2017). Given this bias, we predict that readers would build the structure of a RC, in which the presence of the antecedent of the implicit subject is guaranteed. Despite being more complex, a RC allows immediate resolution of the dependency without further burdening memory.

Furthermore, Korean allows for free word order, and Case and Postpositions can signal the grammatical relations. But, at the same time, grammatical relations are still less easy to understand in noncanonical word orders. Why? One possibility is that case information, even though it’s a purely grammatical feature, is not perfectly maintained. We can probe this idea by testing whether distance affects *scrambled* case. The distance test is critical for examining whether grammatical information is maintained regardless of distance. If grammatical case information decays over time and is retrieved at the gap, the effect of scrambling should be observed at later regions in the long distance conditions.

References

Acuña-Fariña, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua*, *143*, 108-128.

Aelbrecht, L. (2010). *The Syntactic Licensing of Ellipsis*. Amsterdam/New York: John Benjamins.

Aelbrecht, L. (2016). What ellipsis can do for phases and what it can't, but not how. *The Linguistic Review* (pp. 453-482).

Alexiadou, A., & Gengel, K. (2012). NP ellipsis without focus movement/projections: The role of Classifiers. In I. Kučerová & A. Neeleman (Eds.), *Contrasts and Positions in Information Structure* (pp. 177-205). Cambridge: Cambridge University Press.

Alexiadou, A., Haegeman, L., & Stavrou, M. (2008). *Noun phrase in the generative Perspective* (Vol. 71). Walter de Gruyter.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, *51*(1), 23-54.

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.

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.

Bader, M., Bayer, J., Hopf, J. M., & Meng, M. (1996). Case-assignment in processing German verb-final clauses. In *Proceedings of the NELS 26 Sentence Processing*

Workshop, MIT Occasional Papers in Linguistics (Vol. 9, pp. 1-25).

Baker, M. (1988). Theta theory and the syntax of applicatives in Chichewa. *Natural Language & Linguistic Theory*, *6*(3), 353-389.

Baltin, M. (2012). Deletion versus pro-forms: an overly simple dichotomy? *Natural Language & Linguistic Theory*, *30*(2), 381-423.

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.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(7), 1-23.

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 317-333.

Bernstein, J. B. (2001). The DP hypothesis: Identifying clausal properties in the nominal domain. In M. Baltin & C. Collins (Eds.), *The Handbook of Contemporary Syntactic Theory* (pp. 536-561). Blackwell.

Bever, T. G., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry*, *19*, 35-43.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, *34*(6), 774-806.

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological), 26*(2), 211-243.

Bruening, B. (2001). QR obeys superiority: Frozen scope and ACD. *Linguistic inquiry*, *32*(2), 233-273

Brugè, L. (1996). Demonstrative movement in Spanish: A comparative approach. *Working Papers in Linguistics*, *6*(1), 1-53.

Brugè, L. (2002). The positions of demonstratives in the extended nominal projection. *Functional structure in DP and IP: The cartography of syntactic structures*, *1*, 15-53.

Brugè, L., & Giusti, G. (1996). On demonstratives. *GLOW talk, Athens*.

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.

Carlson, G. N. (1987). Same and different: Some consequences for syntax and semantics. *Linguistics and Philosophy*, *10*(4), 531-565.

Chapman, C., & Kučerová, I. (2016). Structural and semantic ambiguity of why-questions: An overlooked case of weak islands in English. *Proceedings of the Linguistic Society of America*, *1*, 15-1.

Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language, 52*(1), 144-169.

Chisholm, M. (2003). Ellipsis in DP. *Unpublished master's thesis, University of California, Santa Cruz*.

Chomsky, N. (1977). On wh-movement. In: P. Culicover, T. Wasow, & A. Akmajian (Eds.). *Formal Syntax*. New York: Academic Press, 71-132.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, Holland Cinnaminson, [N.J.]: Foris Publications.

Chomsky, N. (1986). *Barriers* (Vol. 13). MIT press.

Chomsky, N. (2001). Derivation by Phase. In M. Kenstowicz (Eds.), *Ken Hale: A Life in Language* (pp. 1-52). Cambridge, MA: The MIT Press.

Chomsky, N. (2008). On phases. *Current Studies in Linguistics Series*, *45*, 133.

Chow, W.-Y., & Zhou, Y. (2019). Eye-tracking evidence for active gap-filling regardless of dependency length. *Quarterly Journal of Experimental Psychology*, *72*(6), 1297-1307.

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

Collins, C. (2015). Adjunct Deletion. Ms., NYU.

Collins, C. (2018). Quantifier domain restriction as ellipsis. *Glossa: a journal of general linguistics*, *3*(1), 60.

Corver, N., & van Koppen, M. (2010). Ellipsis in Dutch possessive noun phrases: a micro-comparative approach. *The Journal of Comparative Germanic Linguistics, 13*(2), 99-140.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences*, *14*(4), 597-612.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.), *Natural language parsing* (pp. 94-128). Cambridge: Cambridge University Press.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler Syntax*: Oxford University Press.

Cunnings, I., & Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language, 102*, 16-27.

Davies, W. D., & Dubinsky, S. (2003). On extraction from NPs. *Natural Language & Linguistic Theory*, *21*(1), 1-37.

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

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.

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

Elbourne, P. (2001). E-type anaphora as NP-deletion. *Natural Language Semantics*, *9*(3), 241-288.

Elbourne, P. (2008). The interpretation of pronouns. *Language and Linguistics Compass*, *2*(1), 119-150.

Engdahl, E. (1983). Parasitic Gaps. *Linguistics and Philosophy, 6*(1), 5-34.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and Syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, *47*(2), 250-272.

Fiengo, R., & May, R. (1994). *Indices and identity* (Vol. 24). MIT press.

Fodor, J. D. (1978). Parsing Strategies and Constraints on Transformations. *Linguistic Inquiry, 9*(3), 427-473.

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.

Fox, D. (2000). *Economy and semantic interpretation* (Vol. 35). MIT press.

Fox, D., & Lasnik, H. (2003). Successive-cyclic movement and island repair: The difference between sluicing and VP-ellipsis. *Linguistic inquiry, 34*(1), 143-154.

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. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 129-189). Cambridge, England: Cambridge University Press.

Frazier, L. (2008). Processing ellipsis: A processing solution to the undergeneration problem. In *Proceedings of the 26th west coast conference on formal linguistics* (pp. 21-32). Somerville, MA: Cascadilla Press.

Frazier, L., & Clifton Jr, C. (1989). Successive cyclicity in the grammar and the parser. *Language and cognitive processes*, *4*(2), 93-126.

Frazier, L., & Clifton, Jr, C. (2001). Parsing coordinates and ellipsis: Copy α. *Syntax*, *4*(1), 1-22.

Frazier, L., & Flores D'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*(3), 331-344.

Frazier, L., Munn, A., & Clifton, C. (2000). Processing Coordinate Structures. *Journal of Psycholinguistic Research, 29*(4), 343-370.

Gazdar, G., Klein, E., Pullum, G. K., & Sag, I. A. (1985). *Generalized phrase structure grammar*. Harvard University Press, Cambridge, Mass.

Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1-76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain: Papers from the first mind articulation project symposium.* (pp. 94-126). Cambridge, MA, US: The MIT Press.

Gibson, E., & Thomas, J. (1999). Memory Limitations and Structural Forgetting: The Perception of Complex Ungrammatical Sentences as Grammatical. *Language and Cognitive Processes, 14*(3), 225-248.

Gibson, E., & Warren, T. (2004). Reading-Time Evidence for Intermediate Linguistic Structure in Long-Distance Dependencies. *Syntax, 7*(1), 55-78.

Giusti, G. (1997). The categorial status of determiners. In L. Haegeman (Eds.), *The new comparative* syntax (pp. 95-123). London: Longman.

Giusti, G. (2002). The functional structure of noun phrases: A bare phrase structure approach. *Functional structure in DP and IP: The cartography of syntactic structures, 1*, 54-90.

Gordon, P. C., & Hendrick, R. (1997). Intuitive knowledge of linguistic co-reference. *Cognition, 62*(3), 325-370.

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.

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.

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.

Grinder, J., & Postal, P. M. (1971). Missing antecedents. *Linguistic Inquiry*, *2*, 269-312.

Grodner, D., & Gibson, E. (2005). Consequences of the Serial Nature of Linguistic Input for Sentenial [Sentential] Complexity. *Cognitive Science: A Multidisciplinary*

Journal of Artificial Intelligence, Linguistics, Neuroscience, Philosophy, Psychology, 29(2), 261-290.

Guardiano, C. (2010). Demonstratives and the structure of the DP: crosslinguistic remarks. In *Workshop on Disharmony in Nominals, Linguistics Association of Great Britain Annual Meeting, Leeds University*.

Hankamer, J., & Sag, I. (1976). Deep and Surface Anaphora. *Linguistic Inquiry, 7*(3), 391-428.

Hardt, D. (1999). Dynamic Interpretation of Verb Phrase Ellipsis. *Linguistics and Philosophy, 22*(2), 187-221.

Hartman, J. (2011). The semantic uniformity of traces: Evidence from ellipsis parallelism. *Linguistic Inquiry, 42*(3), 367-388.

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.

Hiraiwa, K. (2005). *Dimensions of symmetry in syntax: Agreement and clausal architecture* (Doctoral dissertation). MIT, Cambridge, MA.

Hornstein, N. (1994). An Argument for Minimalism: The Case of Antecedent-Contained Deletion. *Linguistic Inquiry, 25*(3), 455-480.

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

Huang, C. T. J. (1982). Logical relations in Chinese and the theory of grammar. PhD diss.

Huddleston, R. D. (2002). *The Cambridge grammar of the English language*. Cambridge, U.K. New York: Cambridge University Press.

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.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316-339.

Jayaseelan, K. A. (1990). Incomplete VP deletion and gapping. *Linguistic Analysis, 20*, 64-81.

Johnson, K. (2001). What VP ellipsis can do, and what it can't, but not why. In M. Baltin & C. Collins (Eds.), *The Handbook of Contemporary Syntactic Theory* (pp. 439-479): Blackwell.

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.

Kayne, R. S. (1994). *The antisymmetry of syntax* (Vol. 25). Mit Press.

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.

Kehler, A. (2000). Coherence and the Resolution of Ellipsis. *Linguistics and Philosophy, 23*(6), 533-575.

Keine, S (2015). Locality domains in syntax: Evidence from sentence processing. Ms., University of Massachusetts Amherst.

Kester, E.-P. (1996). Adjectival inflection and the licensing of empty categories in DP1. *Journal of Linguistics, 32*(1), 57-78.

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

Kim, N., Brehm, L., & Yoshida, M. (2019). The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval. *Language, Cognition and Neuroscience*, *34*(2), 190-213.

Kim, N., Brehm, L., Sturt, P., & Yoshida, M. (2019). How long can you hold the filler: maintenance and retrieval. *Language, Cognition and Neuroscience*, 1-26.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580-560.

Ko, H. (2005). Syntax of Why-in-situ : Merge Into [SPEC,CP] in the Overt Syntax. *Natural Language & Linguistic Theory, 23*(4), 867-916.

Kush, D. W. (2013). *Respecting relations: Memory access and antecedent retrieval in Incremental sentence processing* (Doctoral dissertation). University of Maryland,

Kush, D., Lohndal, T., & Sprouse, J. (2018). Investigating variation in island effects. *Natural Language & Linguistic Theory, 36*(3), 743-779.

Kush, D., & Phillips, C. (2014). Local anaphor licensing in an SOV language: implications for retrieval strategies. *Frontiers in Psychology, 5*, 1252.

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.

Larson, R. K. (1988). *Light predicate raising*. Cambridge, MA: Cambridge, MA : Lexicon Project, Center for Cognitive Science, MIT.

Lasnik, H. (1999). Pseudogapping puzzles. Fragments: Studies in ellipsis and gapping, (Eds.), S. Lappin & E. Benmamoun (pp. 141-174): Oxford: Oxford University Press.

Lasnik, H. (2001). When can you save a structure by destroying it? In M. Kim, & U. Strauss (Eds.), *Proceedings of the 31st Annual Meeting of the North East Linguistic Society,* (pp. 301-320), Amherest MA: GLSA.

Lasnik, H. (2005). Review of Jason Merchant, the syntax of silence. *Language, 81*, 259-265.

Lasnik, H., & Park, M. (2013). Locality and MaxElide in extraction out of elided VP. In Y. Miyamoto, D. Takahashi, H. Maki, M. Ochi, K. Sugisaki & A. Uchibori (Eds.), *Deep Insights, Broad Perspectives: Essays in Honor of Mamoru Saito*, (pp. 235-256). Tokyo: Kaitakusha.

Lasnik, H., & Saito, M. (1984). On the nature of proper government. *Linguistic Inquiry 15*:235-289.

Lee, M.W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, *33*(1), 51-73.

Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-Based Forgetting in Verbal Short-Term Memory. *Journal of Memory and Language, 59*(2), 200-222.

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.

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

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.

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.

Lightfoot, D. (1989). The Child's Trigger Experience: Degree-0 Learnability. *Behavioral and Brain Sciences, 12*(2), 321-334.

Llombart-Huesca, A. (2002). Anaphoric One and NP-Ellipsis. *Studia Linguistica, 56*(1), 59-89.

Lobeck, A. C. (1995). *Ellipsis: Functional heads, licensing, and identification*. New York: Oxford University Press.

Lobeck, A. (2007). Ellipsis in DP. In *The Blackwell Companion to Syntax* (pp. 145-173): Blackwell Publishing.

Mahowald, K., Graff, P., Hartman, J., & Gibson, E. (2016). SNAP Judgments: A Small N Acceptability Paradigm (SNAP) for Linguistic Acceptability Judgments. *Language: Journal of the Linguistic Society of America, 92*(3), 619-635.

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.

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.

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: Evidence from Sluicing. *Journal of Memory and Language, 64*(4), 327-343

May, R. (1977). *The Grammar of Quantification*. Doctoral dissertation. MIT, Cambridge, MA.

May, R. (1985). *Logical Form: Its Structure and Derivation*: MIT Press.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*(2), 111-123.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*(3), 817-835.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 46, pp. 155-200). San Diego: Academic Press.

McElree, B., & Bever, T. G. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research, 18*(1), 21-35.

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.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*(1), 67-91.

Merchant, J. (2000). Antecedents-contained Deletion in Negative Polarity Items. *Syntax, 3*(2), 144-150.

Merchant, J. (2001). *The syntax of silence : sluicing, islands, and the theory of ellipsis*. Oxford University Press.

Merchant, J. (2005). Fragments and ellipsis. *Linguistics and Philosophy, 27*(6), 661-738.

Merchant, J. (2008). An Asymmetry in Voice Mismatches in VP-Ellipsis and Pseudogapping. *Linguistic Inquiry, 39*(1), 169-179.

Merchant, J. (2013a). Diagnosing ellipsis. *Diagnosing syntax, 1*, 537-542.

Merchant, J. (2013b). Voice and Ellipsis. *Linguistic Inquiry, 44*(1), 77-108.

Merchant, J. (2014). Gender mismatches under nominal ellipsis. *Lingua, 151*, 9-32.

Miyamoto, E. T., & Takahashi, S. (2004). Filler-gap dependencies in the processing of scrambling in Japanese. *Language and Linguistics*, *5*(1), 153-166.

Munn, A. (1993). *Topics in the syntax and semantics of coordinate structures*. Doctoral dissertation, University of Maryland, College Park.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53*(1), 53-81.

Nakatani, K., & Gibson, E. (2008). Distinguishing theories of syntactic expectation cost in sentence comprehension: evidence from Japanese. *Linguistics*, *46*(1), 63-87.

Nicenboim, B., Logačev, P., Gattei, C., & Vasishth, S. (2016). When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in Psychology*, *7*, 280-280.

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and Confirmatory Analyses in Sentence Processing: A Case Study of Number Interference in German. *Cognitive Science, 42*(54), 1075-1100.

Nicenboim, B., Vasishth, S., Gattei, C., Sigman, M., & Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. *Frontiers in psychology*, *6*, 312.

Nicol, J. L., Fodor, J. D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *20*(5), 1229-1238.

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

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*(1), 5-19.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory.

Journal of Memory and Language, *55*(4), 601-626.

Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in psychology*, *6*, 384.

Panagiotidis, P. (2000). Demonstrative determiners and operators: The case of Greek. *Lingua*, *110*(10), 717-742.

Panagiotidis, P. (2003). Empty nouns. *Natural Language & Linguistic Theory, 21*(2), 381-432.

Park, D. (2016). VP as an ellipsis site in Korean: Evidence for the derivational PF deletion theory. In *Proceedings of North East Linguistic Society* (Vol. 46, pp. 109-122).

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language, 94*, 272-290.

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2017). The cue-based retrieval theory of sentence comprehension: New findings and new challenges. In L. Escobar, V. Torrens, & T. Parodi (Eds.), *Language processing and disorders.* Newcastle: Cambridge Scholars Publishing.

Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology, 69*(5), 950-971.

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.

Pearl, L. & Lidz, J. (2013). Parameters in Language Acquisition. In Grohmann, K. & Boeckx, C. (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.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*(4): 795-823.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Eds.), *Experiments at the Interfaces* (Vol. 37, pp. 147-180). Bingley, UK: Emerald Publications.

Pickering, M., & Barry, G. (1991). Sentence Processing without Empty Categories. *Language and Cognitive Processes, 6*(3), 229-259.

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.

Postal, P. (1969). *Anaphoric islands. Chicago Linguistic Society, 5,* 205-239.

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

Ritter, E. (1991). Two functional categories in noun phrases. *Syntax and semantics, 25*, 37-62.

Roberts, I. (2011). FOFC in DP: Universal 20 and the nature of demonstratives. *Unpublished ms*.

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

Ross, H. (1967). *Constraints on variables in syntax* (Doctoral dissertation, Massachusetts Institute of Technology).

Ross, H. (1984). Inner islands. In C. Brugman, & M.Macaulay (Eds.), *Annual Meeting of the Berkeley Linguistics Society* (pp. 258-65). University of California, Berkeley.

Sag, I. A. (1976). *Deletion and logical form* (Doctoral dissertation, Massachusetts Institute of Technology).

Sag, I. A., & Hankamer, J. (1984). Toward a theory of anaphoric processing. *Linguistics and philosophy*, *7*(3), 325-345.

Schneider, D., & Phillips, C. (2001). Grammatical search and reanalysis. *Journal of Memory and Language*, *45*(2), 308-336.

Schuyler, T. (2001). Wh-movement out of the site of VP ellipsis. *Syntax and semantics at Santa Cruz*, *3*, 1-20.

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.

Sprouse, J., & Almeida, D. (2017). Design Sensitivity and Statistical Power in Acceptability Judgment Experiments. *Glossa: A Journal of General Linguistics, 2*(1), 1-32.

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language*, *88*(1), 82-123.

Staub, A., & Clifton Jr, C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 425-436.

Stepanov, A., & Stateva, P. (2015). Cross-linguistic evidence for memory storage costs in filler-gap dependencies with wh-adjuncts. *Frontiers in psychology*, *6*, 1301.

Stepanov, A., & Tsai, W. T. D. (2008). Cartography and licensing of wh-adjuncts: a cross-Linguistic perspective. *Natural Language & Linguistic Theory*, *26*(3), 589-638.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location.

Language and cognitive processes, *1*(3), 227-245.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*(3), 542-562.

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.

Sturt, P., Pickering, M. J., Scheepers, C., & Crocker, M. W. (2001). The preservation of structure in language comprehension: Is reanalysis the last resort?. *Journal of Memory and Language*, *45*(2), 283-307.

Takahashi, S., & Fox, D. (2005). MaxElide and the re-binding problem. In *Proceedings of SALT* (Vol. 15, pp. 223-240).

Tanenhaus, M. K., Boland, J., Garnsey, S. M., & Carlson, G. N. (1989). Lexical structure in parsing long-distance dependencies. *Journal of psycholinguistic research*, *18*(1), 37-50.

Tanenhaus, M. K., & Carlson, G. N. (1990). Comprehension of Deep and Surface Verbphrase Anaphors. *Language and Cognitive Processes, 5*(4), 257-280.

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.

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.

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language, 48*(4), 740-759.

Travis, L. D. (1984). *Parameters and effects of word order variation* (Doctoral dissertation, Massachusetts Institute of Technology).

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the Processing of Unbounded Dependencies: An Eye-Tracking Study. *Journal of Memory and Language*, *35*(3), 454-475.

Ur, S. (2004). The Form of Semitic Noun Phrase. *Lingua*, *114*(12), 1465-1526.

Valois, D. (1991). *The Internal Syntax of DP, unpublished Ph. D* (Doctoral dissertation, UCLA, Los Angeles).

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.

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.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55*(2), 157-166.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65*(3), 247-263.

Vasishth, S., & Lewis, R. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language, 82*(4), 767-794.

Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PloS one*, *8*(10), e77006.

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.

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.

Wagers, M. W., & Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics*, *45*(2), 395-433.

Wagers, M. W., & Phillips, C. (2014). Going the Distance: Memory and Control Processes in Active Dependency Construction. *Quarterly Journal of Experimental Psychology, 67*(7), 1274-1304.

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119-161). Cambridge, MA: MIT Press.

Warren, T., & Gibson, E. (2002). The Influence of Referential Processing on Sentence Complexity. *Cognition: International Journal of Cognitive Science, 85*(1), 79-112.

Williams, E. (1978). Across-the-Board Rule Application. *Linguistic Inquiry, 9*(1), 31-43.

Xiang, M., Wang, S., & Cui, Y. (2015). Constructing covert dependencies-The case of Mandarin wh-in-situ dependency. *Journal of Memory and Language*, *84*, 139-166.

Yoshida, M. (2010). “Antecedent-contained” sluicing. *Linguistic Inquiry*, *41*(2), 348-356.

Yoshida, M., Nakao, C., & Ortega-Santos, I. (2015). The syntax of Why-Stripping. *Natural Language & Linguistic Theory, 33*(1), 323-370.

Yoshida, M., Potter, D., & Hunter, T. (2018). Condition C reconstruction, clausal ellipsis and island repair. *Natural Language & Linguistic Theory*, 1-30.

Yoshida, M., Wang, H., & Potter, D. (2012). Remarks on “Gapping” in DP. *Linguistic Inquiry, 43*(3), 475- 494.