THE UNIVERSITY OF CHICAGO

THE ENCODING, MAINTENANCE AND RETRIEVAL OF COMPLEX LINGUISTIC REPRESENTATIONS IN WORKING MEMORY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE HUMANITIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF LINGUISTICS

BY
CHI DAT LAM

CHICAGO, ILLINOIS
AUGUST 2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This dissertation is the end-product of a journey, a journey I am blessed to have travelled on with the support of people that I forever hold dear to my heart. It is hard to imagine what it is like without any of them. For that, I am thankful.

First and foremost, I want to thank my advisor, Ming Xiang. I cannot imagine having a better advisor. I often discussed with friends about what makes a great advisor and I always felt wholeheartedly that personally for me, a great advisor has high standards but is also hands-on. Ming is that person. She wants me to achieve the best version of myself by having good experimental techniques, using the most appropriate statistical analysis (thanks for asking me to learn Bayesian stats!) and writing the clearest paragraphs. But she is also willing to spend two or three hours a week with each of her students to talk through ideas, follow up with their progress and also just chat and support her students on a personal level. Working with Ming made me update my own hypothesis about what makes a great advisor: they are also personable. Ming cares for me, and each of her students, as a mentee but also a person. Thank you, Ming, for being my best advisor!

I also want to express my gratitude to the rest of my committee: Allyson Ettinger and Ed Awh. Allyson taught me everything I know about computational linguistics and thanks to her, I felt technologically capable to explore even the trickiest data computationally. I also really enjoy chatting with Allyson, who gave me her computationally-informed perspective on psycholinguistics and also great career advice. Ed made me engrossed in theories of working memory (which are so cool!) and engaged in the literature outside of my little world of sentence comprehension. I wish I have chatted with you more because every time I did, I learned something new and had so many ideas about possible explorations.

I am also indebted to other professors at the university who have mentored me in one way or another. Thanks to Salikoko Mufwene (Sali) for being my first advisor, for guiding my first steps in the field and for teaching me how to think evolutionarily. Thanks to Dan Margoliash for my brief foray into the world of birdsong and for making me ponder about the brain in an ecological perspective. Thanks to Monica Do for interesting discussions about psycholinguistics and for the fun chats when we bumped into each other on the shuttle.

Thanks to Diane Brentari, Alan Yu, Karlos Arregi, Lenore Grenoble, Chris Kennedy, Erik Zyman, Howard Nusbaum, Susan Goldin-Meadow and Ed Vogel for teaching me a lot of what I know about linguistics and psychology. I would also like to express my gratitude towards Julie Tetel Andresen and Edna Andrews, who have ignited my love for linguistics during my time as an undergraduate student.

My dissertation would not have been completed without the experiments and the people involved in them. Thanks to my participants, especially the 55 participants in the EEG experiments, for being patient with my memory-demanding sentences. Thank you, Sam Arnett, Daniel Yue and Carla Escalante, for aiding me in preparing the EEG experiments and also for helping lessen the loneliness of doing experimental work. I also want to thank past and current members of the UChicago Language Processing Lab, for giving me comments and suggestions on my experiments. Special thanks to Jinghua Ou, who, along with Ming, taught me everything about EEG and who was extremely patient with my incessant questions about data analysis. I also appreciate insights on my work from professors and peers who I had the fortune to meet at various conferences and workshops, which include but are not limited to Brian Dillon, Jon Brennan, Jon Sprouse, Tzu-Yun Tung, Steve Schwering, Eun-Kyoung Rosa Lee, Nicolaus Schrum, Margot Berg, Marlyse Baptista, Enoch Aboh.

No one goes through their PhD without supportive friends and I am no exception. My friends in and out of the program supported me unconditionally. For that, I feel very lucky and very thankful. Thank you, Eszter Ronai, for being a second psycholinguistics mentor to me, for always supporting me with the best encouragements, for co-coordinating LEAP with me and for way too many fond memories to recount. Thank you, Sanghee Kim, for "peer-coding" with me, for lightening the deep dark dungeon of our Landahl office and for the most thoughtful and thought-provoking conversations about psycholinguistics and also about life. Thank you, Aurora Martinez del Rio, for always joining me on cafe work sessions (without which I probably could not finish this dissertation), for sending me various important information and reminding me about deadlines and for reminding me every time that I got this. Thank you, Han Nguyen, for also joining me on cafe work sessions and for always being there for me through the important milestones of my life. Also, quick detour to thank the three cafes that supported the caffeine intake and the perfect atmosphere for me to write my dissertation: Sweet Bean, Cafe Press and Dollop. Thank you, Matt Hewett, for various fun hangouts and also fun debates on the nature of human languages. Thanks also to many other friends at the University of Chicago who have made my time there so worthwhile and so special, which include Omer Eren, Sam Gray, Laura Stigliano, Mike Tabatowski, Lucas Fagen, Emre Hakguder, Adam Singerman, Stephanie Reyes, Darragh Winkelman, Kate Henninger, Corinne Kasper, Parker Robbins, Zeineb Sellami, Kutay Serova, Pamela Sugrue, Brianna Wilson, Akshay Aitha, Jonathan WuWong, Thomas Sostarics, Weijie Xu, Serpil Karabuklu, Jessica Kantarovich, Nelson Medina, Claire Bergey, James Waller, Sunny Lee, Casey Ferreira, Andy Savoy, Anastassia Sorokina, Gaby Nicole, Mia Nicole and Carlos Cardenas, Jordan Greer. Of course, I can never forget support and love from friends from other avenues of life who constantly encourage and cheer me on, including among others: Thu Nguyen, Ha Nguyen, Nisakorn Valyasevi, Thao Duong, Vu Lan Nguyen, Ilya Raskin, Drew Heckman, Silvie Senauke, Elliot Karl, Juan Garcia, Michael Mendelsohn, Jasmine Tsai and An Nguyen.

There is always the a special place in my heart for family. Thank you to my cat, Fiora, for the unconditional love and for the precious company during my late nights. Thank you to the Wedell family, especially Carol and Mark, for always making me feel at home. Thank you to my extended family, my aunts, uncles and cousins, for always supporting me, treating me delicious Vietnamese food and showering me with love and care. Special thanks go to my youngest aunt, who calls me once every two weeks just to chat. Talking with her always leaves me feeling warm and loved. And then, of course, thank you to my parents and my two brothers. My mom constantly reminds me to take care of my health and happiness but also celebrates every special moment with me. My dad, an academic, has inspired me to pursue my PhD and keeps giving me great advice during this journey. My middle brother Minh checks in on me and gives me much needed support during the toughest of times. My youngest brother Nghia never fails to make me smile during our weekly call with his stories and encouragements.

Finally, thank you, Justin. You came into my life and changed it forever, for the better.

You know this work inside out, having heard me go on and on about it and read many of my abstracts. You talk about my work to other people much better than I ever can, with such confidence that I envy. Thank you for always being by my side and making my day much better and happier, and of course for always loving me. I am eternally grateful to have you in this journey and in my life.

This material is based upon work supported by the National Science Foundation under Grant No. #BCS-2116952.

ABSTRACT

In everyday life, humans rely on working memory (WM) processes to make sense of relationships between linguistic elements that are not linearly adjacent. For example, to understand the sentence *The dog that the cat chased is cute*, we encode the referent *the dog* into WM, maintain and retrieve it after reading the verb *chased* in order to interpret *the dog* as the object of *chased*. Increased grammatical complexity of the sentence has been shown to increase the effort needed to retrieve the target representation (Gibson, 1998; Lewis and Vasishth, 2005). In this dissertation, I focus on how changing the properties of the target representation - its complexity and internal coherence - affects the effort needed for the three WM processes, especially the under-studied maintenance process. Understanding this interaction can help inform our understanding of how humans package and store linguistic units in WM and better characterize WM architecture during sentence comprehension. To give a multi-dimensional view of the impact on WM processes, I approached the question from both behavioral and neural perspectives. Furthermore, through EEG experiments, I examined the role of the Sustained Anterior Negativity (SAN) in maintenance during sentence comprehension.

Chapters 2 and 3 investigate how encoding, maintenance and retrieval of complex noun phrase (NP) representations differ from those of their simple counterparts through both a behavioral and neural lens, respectively. Through self-paced reading experiments in Chapter 2, I demonstrate that while the effects of representational complexity on encoding and retrieval were inconsistent, having a complex target NP always facilitated maintenance. This facilitation results from two factors: increased distinctiveness of the complex representations leading to decreased vulnerability to interference, and increased activation level due to increased encoding effort. Chapter 3 describes an EEG experiment that followed up the behavioral experiments in Chapter 2. Previous works (Fiebach et al., 2002; King and Kutas, 1995; Phillips et al., 2005) found the SAN when there was increased processing cost during maintenance. In this EEG experiment, the sentences with simple NPs elicited a sustained global negativity relative to sentences with complex NPs during maintenance, giving evidence that increased representational complexity of the target NP decreases maintenance

effort.

Chapter 4 investigates how the encoding, maintenance and retrieval of internally coherent NP representations differ from those of their incoherent counterparts using both behavioral and EEG experiments. Additionally, these experiments further explore the role of representational complexity and how it interacts with internal coherence. Two self-paced reading experiments revealed that having a coherent target NP facilitates encoding, early maintenance and to a lesser extent, retrieval. They also provided evidence for less effort during early maintenance of complex target NPs. Accordingly, in the EEG experiment, the conditions with simple or incoherent target NPs triggered a SAN during early maintenance relative to the conditions with complex or coherent target NPs. All these experiments together suggest that having coherent features eases both encoding and early maintenance processes.

Overall, the findings demonstrate that increasing complexity or coherence at the representation level consistently facilitates WM maintenance. They also show that the SAN provides a viable ERP index for processing effort during maintenance though its behavior might be modulated by predictive processes.

CHAPTER 1

GENERAL INTRODUCTION

In everyday life, speaking and signing humans comprehend and produce sentences that vary widely in complexity, both in structure (syntactically) and in meaning (semantically). In many of the sentences, we have to process relationships between words that are not adjacent in order to understand or produce the sentence. For example, consider the following sentences:

(1) The person who cuddled the cat loves food.

(2) It is difficult to find Waldo without a magnifying glass.

In sentence (1), though the noun phrase (NP) *the cat* is linearly adjacent to the verb phrase (VP) *loves food*, every English speaker understands that it is the person, not the cat, that loves food. Sentence (2), for some (or even many) English speakers, is an ambiguous sentence. It might be the case that it is hard to find an instance of Waldo that does not have a magnifying glass. Alternatively, some people can interpret the sentence as saying that people without a magnifying glass might struggle to find Waldo. For the latter interpretation, we once again see that while the NP *Waldo* is linearly adjacent to the prepositional phrase (PP) *without a magnifying glass*, English speakers can interpret the sentence without relying on linear adjacency. These relationships between words that go beyond linear adjacency, such as in relative clauses and pronoun resolution, are called *unbounded dependencies*. The processing of unbounded dependencies in human languages partially relies on the use of WM to encode the linguistic element, maintain it and retrieve it when it is necessary for sentence structure building and/or interpretation. For example, in sentence (1), comprehenders would encode the NP *the person* into WM, maintain it through *who cuddled the cat* and retrieve it upon reading the verb *loves* to incorporate *the person* as the subject of *loves* (they would also retrieve the NP upon encountering *cuddled* to incorporate it as the subject of *cuddled*). In this dissertation, I examine how the representation of the target NP - more specifically, its complexity and internal coherence - in unbounded dependencies influences the three key WM processes of encoding, maintenance and retrieval through both behavioral and neural experiments. The examination of this interaction will help contribute to better understanding of the architecture of WM in sentence comprehension.

In this chapter, I will start by giving an overview of what we know about WM in general in section 1.1. I will summarize key models of WM architecture, mainly from findings about visual and verbal WM, as well as understanding about the three WM processes (encoding, maintenance and retrieval) and the two possible mechanisms behind forgetting (decay and interference). In section 1.2, I will describe the debate over the significance of persistent neural activity in WM maintenance. These two sections provide the necessary background to better appreciate models of WM in sentence comprehension and the discussion over the functional significance of the sustained anterior negativity - the prominent persistent EEG signal in sentence comprehension. Both these topics will be elaborated in section 1.3. In section 1.4, I will define representational complexity and internal coherence (and the related concept of similarity). I will also describe what we know about how changing these properties of linguistic representations can affect the WM processes in sentence comprehension. Section 1.5 is a brief note on how the terminologies of *encoding*, *maintenance* and *retrieval* are used in this dissertation. Lastly, section 1.6 provides an outline of the subsequent four chapters of the dissertation.

1.1 Working Memory - Architecture and Processes

1.1.1 Models of Working Memory and Working Memory Capacity

In order to appreciate the three WM processes, it is crucial to understand how WM researchers have characterized WM architecture (an excellent review can be found in Jonides et al. (2008)). One of the earliest models of human WM came from the works of Alan Baddeley and colleagues (Baddeley, 2000, 1966; Baddeley and Hitch, 1974). They proposed a multi-store model of WM where there are separate storage for verbal and visuospatial WM: the phonological loop and the visuospatial sketchpad, respectively. A key character of the phonological loop is that information held in there can be rehearsed while being maintained. In order to facilitate the processing of multimodal stimuli, Baddeley (2000) also added an episodic buffer that connects the phonological loop and the visuospatial sketchpad with each other, with other buffers and with long-term memory (LTM). As such, an important feature of this model is the dissociation between WM and LTM, as well as dissociation between WM involving different modalities.

Some more recent models of WM - called unitary-store models - proposed that WM and LTM are not architecturally separate systems (Cowan, 1999; McElree, 2001; Oberauer, 2002). For example, Cowan (1999)'s embedded-processes model hypothesized that there is a focus of attention that is capable of holding a certain number of representations (the capacity limit - which is 4 ± 1 for this model). Besides representations in this focus of attention, there are also other "activated" representations in LTM that can be accessed into the focus of attention. These activated LTM representations have different activation levels which can depend on many factors, such as recency. They are also more susceptible to forgetting, by decay and/or interference. Thus, while representations in the focus of attention are limited by the capacity limit, activated representations in LTM are limited by mechanisms that lead to forgetting.

While the various unitary-store models all posit a tight connection between WM (or focus of attention) and LTM, some of the specifics vary between models. A contentious property is the capacity limit of the focus of attention. As cited above, Cowan (2010)'s model assumed a capacity limit of 4 ± 1. The model proposed by McElree (2001), however, has a much smaller capacity limit of 1. This much stricter capacity limit resulted from experiments using the speed-accuracy tradeoff paradigm where McElree (1998) observed that while memory accuracies differed between the different items, retrieval speed for all items, except for the most recent item, were similar and the most recent item was retrieved much faster than others. This observation prompted McElree (2001) to give a privileged status to only one item. Cowan (2010)'s capacity limit of 4 also received different types of supporting evidence, both behaviorally and neurally. Many behavioral experiments such as Luck and Vogel (1997) showed that accuracy data are consistent with a capacity limit of 3-4. Additionally, EEG experiments like Vogel and Machizawa (2004) demonstrated an EEG index, the contralateral delay activity (CDA), whose magnitude correlated with the number of items being maintained and asymptoted at the capacity limit of 3-4. To reconcile these different results, Oberauer (2002) proposed a model comprising of 3-4 representations that can be accessed directly but only one of them is in the current focus of attention. There are other accounts of WM architecture that reject the idea of a capacity limit on the number and instead view WM resources as continuous (Bays and Husain, 2008; Wilken and Ma, 2004). In these models, we can theoretically store as many items as we want although precision for each item will decrease with increasing number of items. Given the categorical nature of linguistic representations during sentence comprehension (unlike representations with more continuous qualities such as color or spatial position), it is difficult to entertain these flexible-resource models in sentence processing. Thus, this dissertation will largely incorporate the insights from fixed-capacity models (Cowan, 1999; McElree, 2001; Oberauer, 2002).

1.1.2 Working Memory Processes: Encoding, Maintenance and Retrieval

WM, no matter the domain, depends on three important processes: 1) **encoding** the perceptual representation into the focus of attention or WM, 2) **maintenance** of the representation in the focus of attention or WM and 3) **retrieval** of the representation to the focus. **Encoding** is the process where representations, after perceptual processing, are placed into the focus of attention. However, as the name *focus of attention* suggests, this process is highly modulated by attention and not everything arrives at the focus of attention after perceptual processing. This is how we only hold important information in memory instead of everything processed by our perceptual systems. Since the focus of attention is limited, encoding of new item often displaces an existing item out of the focus if the capacity is reached. In McElree (2001)'s model where the capacity limit is 1 item, encoding any new item requires the displacement of the previously encoded item. In other models with a larger capacity limit (Cowan, 1999; Oberauer, 2002), encoding of new representations into the focus can lead to graded similarity-based interference (Nairne, 2002). When the new representation shares features with existing representations in the focus/WM, these representations compete for those features, leading to *feature overwriting* and degradation of one or both of the representations.

Maintenance is the process where the representation is actively stored in the focus of attention/WM without the presence of the perceptual input while being shielded from some sources of interference such as irrelevant stimuli (Postle, 2006). Similar to the encoding process, this process is also highly modulated by attention, which controls which representations can stay in the focus/WM and which representations are relegated to activated LTM. There are modulatory processes during maintenance that make representations less susceptible from incoming stimuli (Postle, 2006), which gives the name of *working memory*. Neurally, there is much evidence that demonstrated the presence of sustained signals during maintenance through single-cell recordings, fMRI and EEG, possibly reflecting the storage of representations. We will discuss more about sustained signals in the next section, section 1.2.

Retrieval is the process where representations needed for processing are brought into the focus of attention/WM from LTM/activated LTM. It has been shown that this is a parallel content-addressable process instead of a serial scanning process. Recall that McElree (1998)'s model relied on the observation that except for the most recent item, other items have similar rate of access (McElree and Dosher, 1989). This is taken as evidence for a parallel content-addressable process for representations outside of the focus of attention. Most WM models are not explicit about the mechanism of the retrieval process (Jonides et al., 2008), with the exception of models in sentence processing, where most focus is on the retrieval process, as we will see in section 1.3.

1.1.3 Decay and interference

It would be amiss to discuss any memory processes without discussing forgetting. There are two (possibly even one) mechanism for forgetting from WM: **decay** and **interference**. **Decay** is the process whereby information in WM gradually erodes over time, resulting in lower activation level and subsequently, more difficulty to retrieve. The evidence for decay came from observations that recall from WM gets worse over time and that sustained signals during maintenance decline in magnitude over time (Fuster, 1995; Jha and McCarthy, 2000). However, the existence of decay remains controversial since many theorists (Lewandowsky et al., 2009; Nairne, 2002) argued that phenomena attributed to time-based decay can be instead explained by other factors such as interference. Nevertheless, there are still some proposals for neural mechanisms of decay that are, again, compatible with interference.

Unlike decay, interference remains a compelling, albeit complex, mechanism behind forgetting. In the WM literature, researchers often distinguish between proactive interference (interference of the incoming representation from previous representations) and retroactive interference (interference of previous representations from the incoming representation). Interference can also happen during any of the three stages of WM. A classic study of proactive interference came from Wickens (1970). In this experiment, participants memorized categorized lists of three words, then performed another task during the retention period and finally recalled the list. The first three lists always came from the same categories (for example, all nine words of the three lists are flowers). They observed that recall became progressively worse across the lists. The critical trial was the fourth list, which might be of the same category as the first three or of a different category. If the list was from a different category, performance increased to even rival that of the first list, showing lesser extent of retrieval interference. Additionally, even subtle differences (flowers to wild flowers) could trigger this effect if participants were told beforehand that a shift would happen. Gardiner et al. (1972) showed that this could also be the case if the shift warning happened right before recall, hence demonstrating that this effect must be partially caused by the selection process during retrieval. However, there are also proposals for interference as a result of the encoding process (Nairne, 2002; Oberauer and Kliegl, 2006), where representations in WM that share features (such as color, shape, ...) and thus compete on those “feature units", resulting in the decrease of activation level of these representations. Despite the complex empirical landscape behind interference, it is an important cornerstone of most WM models. WM models in sentence comprehension are no exception, as we will see in section 1.3.

1.2 Working Memory and Persistent Neural Activity

Since this dissertation focuses most heavily on the WM maintenance process and also the ERP signature commonly associated with WM maintenance during sentence processing, the sustained anterior negativity (SAN), a review of sustained signals related to WM and objections to their significance is crucial. The association between WM maintenance and persistent neural activity is very prevalent in the WM literature (though most prolific in the visual WM literature) and until recently, remains rather uncontroversial. The intuition behind this association is since people are retaining these representations in WM, there must be neural activity associated with them that persists during the maintenance period. The first examination of persistent neural activity came from Fuster and Alexander (1971), where they performed single cell recording in the prefrontal cortex (PFC) of macaques performing a WM task. They observed that neural firing in the PFC increased the most when the stimulus was presented and when the delay period started. Critically, the firing rate was also higher during the delay period than during the pre-trial period. Funahashi et al. (1989) confirmed these results with a similar experiment using an oculomotor task. In this task, macaques were trained to fixate to a central position during stimulus presentation and a delay of 1-6 s. After the delay period, macaques had to make a saccade into the position of the cue (out of eight possible positions) they saw during stimulus presentation. Out of the 288 neurons they recorded, around 1/3 showed increased or decreased firing during the delay period. Most of these neurons only changed their firing for specific cue locations, showing that this persistent activity is somewhat sensitive to the nature of the representation. Additionally, they found that the magnitude of the sustained activity decreased before an error, hinting that the magnitude of the persistent signal correlates with the fidelity of the representation and thus, the performance on individual trials.

While the pioneering works on persistent neural activity were from single cell recordings in macaques, there have also been successes in identifying sustained signals in humans through both neuroimaging and electrophysiology. PET and fMRI experiments on humans doing WM tasks did not straightforwardly show evidence for sustained signals in PFC like experiments on macaques. For example, Leung et al. (2002) showed PFC activity but only when the memorandum was increased to five items. Other studies such as Curtis and D'Esposito (2006); Srimal and Curtis (2008), among others, did not find persistent activity in dorsolateral PFC like in macaques but in other brain regions such as the superior spur of the precentral sulcus and the posterior intraparietal sulcus. However, attempts to decode WM content from fMRI results using machine learning and not just neural activity averaging were more successful. Both Harrison and Tong (2009) and Serences et al. (2009) managed to decode remembered information from retention period activity and at the same time, showed that information that was not remembered or was discarded could not be decoded from fMRI activity. This pattern proved that the decoding success came from WM-related content and not spillover activity from the perception process, especially since much of the information was decoded from activity in early sensory regions. Subsequent works used machine learning to decode various types of visual features from fMRI data of the delay period in various regions in the occipital, parietal and frontal lobes. Interestingly, regions where WM contents could be decoded and regions that showed sustained delay-period activity often existed in a complementary manner (Postle, 2015). For example, Riggall and Postle (2012) could decode stimulus direction from signals in the visual cortex but did not observe any sustained signal in this region. Conversely, they showed increased activity throughout delay period in parts of the frontal cortex but could not decode features of the stimulus from these areas. Another study by Emrich et al. (2013) found sustained retention-period signals in the frontal and parietal cortex that increased in magnitude when WM load increased. However, the same increase in activity was not present in the early visual regions, where the contents of the stimuli could be decoded. These studies were taken as evidence for two types of signals during WM maintenance: one from the frontal and parietal cortices that reflects the number of representations and not the content of the representations and one from the early sensory regions that reflects the content and not the load. This dual representation is consistent of the idea of content-independent *object files*, *pointers* or *indexicals* in the WM literature (Awh et al., 2007; Kahneman et al., 1992; Thyer et al., 2022; Xu and Chun, 2007), which posited two stages - an earlier stage where objects are individuated and a later stage where the detailed content of the objects is made out.

On the electrophysiological side, scientists have observed many sustained ERP indices that occurred during the delay period. In the visual WM literature, Vogel and Machizawa (2004) observed a slow wave they named the Contralateral Delay Activity (CDA). In this setup (panel A of Fig 1.1), participants saw a bilateral display with the same number of visual objects on each hemifield and were instructed to memorize the objects in one hemifield via a cue (before stimulus presentation). After a 900-ms delay period with only a fixation cross on the screen, participants saw a test array and had to respond whether the hemifield they were cued to memorize had changed or not. Vogel and Machizawa (2004) obtained EEG measures from posterior electrodes from the hemisphere contralateral to the cued hemifield and the ipsilateral hemisphere. They detected a negative slow-wave in the contralateral hemisphere relative to the ipsilateral hemisphere that lasted for the whole delay period, which they called the CDA. Critically, the magnitude of the CDA correlated with the number of objects participants were instructed to memorize and but stopped increasing when WM load exceeded 3-4 objects - the capacity limit as proposed by Cowan (2010). Notably, the magnitude increase between 2 and 4 objects also correlated with individual capacity differences (Unsworth et al., 2015; Vogel and Machizawa, 2004). Another ERP index in the visual WM literature that does not depend on subtraction between hemispheres is the negative slow wave (NSW) (Diaz et al., 2021; Fukuda et al., 2015), which was also observed in posterior electrodes. Similar to the CDA, EEG signals from these electrodes got more negative with increasing WM load and also asymptoted around 3-4 objects. Oscillatory activity in the alpha band within these electrodes also showed similar load-dependent patterns (Fukuda et al., 2015), although Diaz et al. (2021) showed evidence that sustained signals like the CDA and the NSW tracked the number of representations in WM while alpha activity was more sensitive to the content of the representations. This pattern mirrors the aforementioned dual neural codes from fMRI experiments and again provides support for the existence of object files (Awh et al., 2007; Kahneman et al., 1992; Thyer et al., 2022; Xu and Chun, 2007).

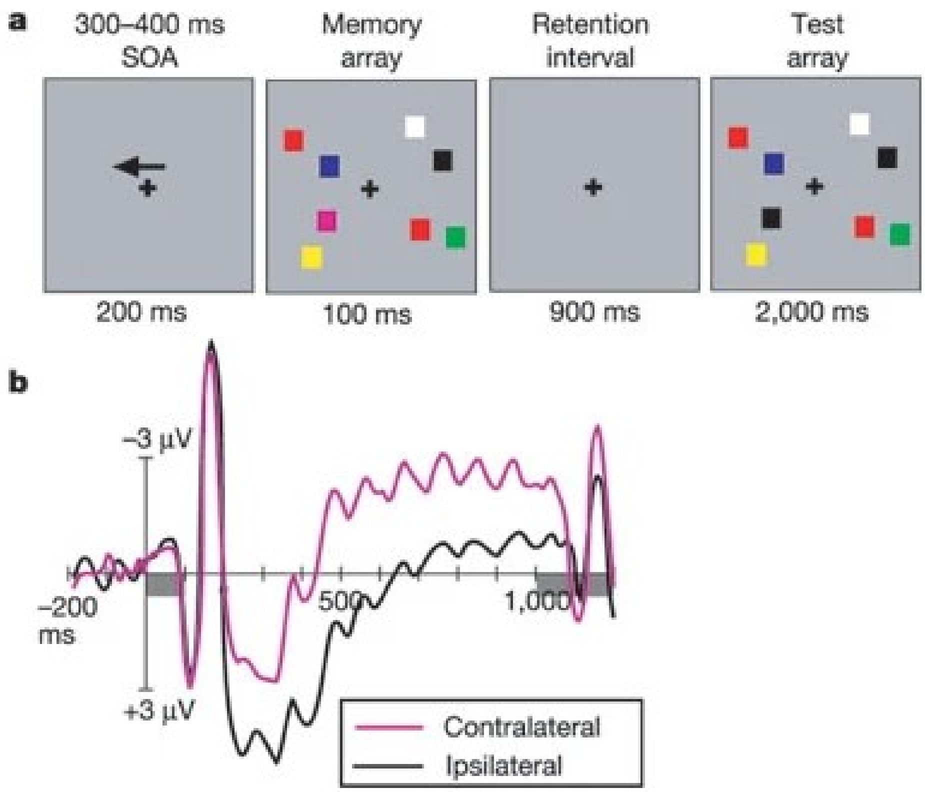

Figure 1.1: The change detection task (a) and the contralateral delay activity (b). SOURCE: Vogel and Machizawa (2004). Reproduced with permission from *Springer Nature*.

Sustained ERP signal during WM maintenance is not only restricted to visual WM. Nolden et al. (2013) and Lefebvre et al. (2013) also detected a sustained signal for auditory WM that is similar to the NSW and the CDA: the auditory sustained anterior negativity (SAN, called auditory SAN here to avoid confusion with the ERP index during sentence comprehension with the same name which will be reviewed in section 1.3). In the experimental condition, participants listened a sequence of 1, 2 or 3 tones differing in timbre or pitch and after a 2000-ms delay period, they listened to another sequence of the same number of tones and responded whether the two sequences were the same or different. In the control condition, the presentation was similar except that participants were instructed to ignore the first sequence and then had to respond whether the final tone of the second sequence was clear or noisy. They found a set size-sensitive slow wave in frontal electrodes - the auditory

SAN - during the delay period that increased in magnitude with increased set size for the experimental conditions. This pattern was not observed for the control conditions, showing that it did not reflect lingering signal from early sensory processes (which are shared between experimental and control conditions). As a result, the auditory SAN has been viewed as the ERP marker for the maintenance of auditory representations.

However, there are several challenges to the significance of persistent neural activity in maintenance of representations in WM (Curtis and Sprague, 2021), the most relevant of which is the idea of "activity-silent" representations. Lewis-Peacock et al. (2012) presented participants with a WM-task where they were instructed to remember two items (a word and a shape, for example). During the retention period, participants were presented with a cue to one of the two items, which indicated that that item would be tested in the initial test. However, a latter test still had a chance of testing the uncued item, requiring participants to still maintain the uncued item in memory. They found that while the the content of the cued item could be decoded from the fMRI data during the retention period, the content of the uncued item could not. Furthermore, in switch trials where the uncued item was later cued, they found that decoding evidence for the previously uncued item increased again. Thus, Lewis-Peacock et al. (2012) argued that sustained activity reflected attention to certain items instead of maintenance of representations of WM since their experiment showed that only the attended items could be decoded, not all the items stored in WM. Similarly, other experiments have shown that decoding performance coild drop gradually but then increased again or was “reactivated" with the presentation of a stimulus (visual or neural) or task instruction (Rose et al., 2016; Sprague et al., 2016; Wolff et al., 2015,1). This again showed that the presence of sustained signals is not necessary for maintenance of WM representations. Supporters of the activity-silent WM representation model have proposed that changes in synaptic weight, in lieu of persistent changes in neuronal firing, as a possible mechanism for storage of WM representations (Stokes, 2015).

There are several responses to the activity-silent WM representation model. Firstly, it is possible that sustained activity is present before reactivation or for unattended stimuli but it could be picked up by the decoding algorithm. Christophel et al. (2018) showed some evidence that irrelevant information was present when more sensitive analysis tools were used with larger sample size. It is also possible that while information about unattended items are not present in fMRI data decoding or electrophysiological data, it may exist elsewhere, such as in oscillatory information (Foster et al., 2016; LaRocque et al., 2013). Foster et al. (2019) also argued that it is possible that information which does not show up as sustained signals has been moved to activated LTM since retrieving unattended items bears many similarities to retrieving items from LTM. This, coupled with the tight connection between sustained signals and some hallmarks of WM (such as their WM load sensitivity), might point to a distinction between “online" information in WM that is reflected by persistent neural activity and “offline" information in activated LTM that is not. Both arguments for and against the central role of persistent neural activity in WM maintenance are relevant to theorists who study WM in sentence comprehension, especially from a neural perspective.

1.3 Working Memory and Sentence Comprehension

Theories about human WM that come from other domains (mainly visual and non-sentence verbal WM) have inspired models of WM in sentence comprehension though the adaptation into sentence processing proved relatively challenging for several reasons. Firstly, as aforementioned in section 1.1, it is still controversial whether different domains rely on different neural substrates and mechanisms for WM. Thus, it is unclear what is shared and not shared between WM in sentence comprehension and WM in other domains. Secondly, human languages are highly hierarchical, making it difficult to determine what constitutes an “item" in WM. For example, does the human mind store *the cat and the dog* as two separate items or as a single item? The difficulty of the determination of the “item" makes it less straightforward to incorporate and test concepts such as capacity limit or WM load. Nevertheless, there are several proposals regarding the role of WM in sentence comprehension which largely divide into capacity-based theories (Gibson, 1998, 2000; Just and Carpenter, 1992) and interference-based theories (Gordon et al., 2001; Lewis and Vasishth, 2005; Van Dyke and McElree, 2006).

1.3.1 Capacity-based Theories

Capacity-based theories attempted to explain psycholinguistic phenomena by relying on the idea of the capacity limit. When there is processing difficulty, capacity-based theories often invoke the trade-off between maintenance and ease of processing. Thus, they propose that maintenance of more materials in WM means more resources are used for storage and fewer are used for processing, resulting in reading slowdown or decreased accuracy.

For example, in Dependency Locality Theory (DLT) (Gibson, 1998, 2000), WM cost is divided into storage cost and integration cost. Storage cost involves the storage of predicted linguistic heads that are needed to fulfill incomplete dependencies. Take the classic example of the subject relative clause (RC) advantage in English:

(3) **Subject RC:** The lawyer who *criticized* the witness admitted the error.

(4) **Object RC:** The lawyer who the witness *criticized* admitted the error.

It is reported that people often have difficulty processing the word *criticized* and subsequent words in sentence (4) with an object RC, relative to sentence (3) with a subject RC. For DLT, a possible explanation is that sentence (4) has higher storage cost than sentence (3) up to the point of retrieval, the verb *criticized*. For example, at the verb *criticized* in (3), only two predicted heads are needed: the object noun for *criticized* and the matrix verb for *the lawyer*. At the same region in (4), the determiner *the*, three predicted heads are needed: the RC subject noun, the RC verb and the matrix verb. Due to the higher number of predicted heads, (4) suffers from higher storage cost up to the verb *criticized*, resulting in fewer resources for retrieval and integration at *criticized* and consequently, processing difficulty relative to (3).

Integration cost is incurred when the incoming word is linked to the current syntactic structure. The more intervening referents needed to be integrated, more resources are dedicated to their integrations and the higher integration cost is required to reactivate the previously built structure to integrate the incoming word. In (3), since the verb *criticized* immediately follows the complementizer *who*, there is no intervening referent within the dependency. In contrast, in (4), the referent *the witness* intervenes between *who* and *criticized* and consmes some resources for its own integration, resulting in decay of the referent *the lawyer*. This resulted in more effort needed to retrieve *the lawyer* and integrate upon reading the verb *critized*. Thus, at the center of DLT is the constant resource trade-off between WM storage and processing, which Gibson (2000) used to explain not only the subject RC advantage but also other psycholinguistic phenomena such as ambiguity effects or the particular difficulty of center-embedded constructions.

Additionally, to support capacity-based theories, there have been studies that linked individual differences in complex span with processing difficulty (Gibson, 1998; Just and Carpenter, 1992; King and Just, 1991). For example, King and Just (1991) compared the subject RC advantage between groups with different spans. They found that participants with low span had more difficulty with sentences with object RC relative to sentences with subject RC, relative to participants with high span. This result demonstrated the connection between individual differences in WM capacity and processing effort, hence providing evidence for capacity-based theories.

1.3.2 Interference-based Theories

Interference-based theories (Lewis and Vasishth, 2005; Lewis et al., 2006) have become the prevailing theories in psycholinguistics since the pioneering Lewis-Vasishth model (abbreviated as LV05) (Lewis and Vasishth, 2005). The model was adapted from the ACT-R model (Anderson, 1996). Like DLT, the LV05 model placed emphasis on the role of intervening referents. Unlike DLT, the LV05 model focused less on the number of referents and more on these referents' content. They also explained psycholinguistic phenomena through interference effects at retrieval (and encoding) and not through storage cost at maintenance. In this theory, referents are stored as bundles of syntactic and semantic features in WM. They are then retrieved in a content-addressable manner through cues at the retrieval site. Processing difficulty was explained by the interference between the target referent and competing referents sharing features, making the target referent harder to be retrieved.

For example, to explain the RC processing assymmetry in sentences (3) and (4) above, the LV05 model reasoned that the intervening referent in (4) *the witness* shared both *+subject* and *+human* features with the target referent *the lawyer* since it is both the subject of the RC and refers to a human entity. This is unlike in (3) where there is no intervening referent. At the retrieval site *criticized* in (4), the verb is looking for an object NP that is typically human. Since *the lawyer* and *the witness* share the *+human* feature which matches the cue, *cue overload* (Van Dyke and McElree, 2006) ensues, leading to retrieval difficulty in (4). What type of cues are relevant during cue-based retrieval is still under investigation but a recent iteration of the model (Engelmann et al., 2019) accommodated both syntactic (*+subject*) and semantic cues (*+animate*). Lexico-semantic cues such as *+shatterable* might also be relevant, as shown by an illustration of "illusion of plausibility" (Cunnings and Sturt, 2018). There has also been success in modeling similarity between word embeddings as cue similarity (Smith and Vasishth, 2020).

While the RC assymetry example above demonstrated the workings of the LV05 model, it cannot distinguish between LV05 model and capacity-based models like DLT since sentences with ORC have one intervening referent while sentences with SRC do not. Nevertheless, there are studies that investigated similarity-based interference while keeping the number of competing referents constant. Gordon et al. (2001) showed that the ORC slowdown was significantly diminished if the intervening NP was a pronoun instead of a noun (with its determiner). Interference can also occur proactively when the competing referents are not in the sentence itself. In a dual task paradigm, Van Dyke and McElree (2006) instructed participants to memorize a word list such as *table-stove-truck* and then read a sentence while holding the word list in mind for a subsequent task. The sentence they read was one of the sentences below:

(5) It was the boat that the guy who lived by the sea **sailed**.

(6) It was the boat that the guy who lived by the sea **fixed**.

At the verb *sailed* in (5), one of the cues for the object NP to be retrieved is *+sailable* and none of the objects in the word list are sailable. To the contrary, in (6), one of the retrieval cues is *+fixable* and all of the objects in the word list can be fixed and thus have the feature *+fixable*. Cue-based retrieval would predict that participants have more difficulty processing the verb in (6) than in (5) due to interference from words in the word list having a feature targeted by the cue. Indeed, Van Dyke and McElree (2006) detected a slower RT at the verb in (6).

The LV05 model, while placing less emphasis on the maintenance process, still factors in decay in their model. As more time passes since the most recent retrieval or encoding of the referent, its activation level slowly decreases due to decay. Thus, the model handles decay in terms of temporal passage, not number of intervening referents that need to be integrated. Nevertheless, decay as a factor was rarely discussed in the sentence comprehension literature using the LV05 model as the model assigned more importance to interference. Van Dyke and Lewis (2003) manipulated both the distance between the encoding and retrieval sites and the degree of similarity between the target referent and the intervening referent. They discovered that whereas increased similarity between the referents penalized both RT and grammaticality judgments, while increased distance did not. They argued that the reactivations at other retrieval sites constantly increased the activation level of the referent, making decay less central. Thus, decay might play a bigger role if there are fewer or no reactivations in the intervening materials. Alternatively, as discussed in section 1.1, it is possible that decay is just a manifestation of increased interference instead of a distinct process.

While the LV05 model explained interference at the retrieval site as a result of *cue overload* at retrieval, other studies have pointed out that encoding interference can also explain interference at retrieval. Proponents of encoding interference (Gordon et al., 2001; Villata et al., 2018) rely on *feature overwriting* (Nairne, 1990, 2002; Oberauer and Kliegl, 2006) to account for interference effects. When a new referent that shares a subset of features with the target referent is encoded into WM, it will compete with the target referent for those features, resulting in degradation of the representation of the target referent, lowering its activation level and making it more difficult to retrieve the target referent. Rich and Wagers (2020) and Kim and Xiang (2023) found that inhibitory interference effects could still be detected when the shared feature was not cued, providing evidence for encoding interference. On the other hand, results from Van Dyke and McElree (2006) shown above could not be explained by encoding interference since only the retrieval sites differed between the conditions. Since encoding and retrieval interference are not mutually exclusive, many authors (Mertzen et al., 2023; Villata et al., 2018) have suggested that the interference patterns they observed could be the result of both processes at work.

In a nutshell, interference-based theories emphasize the content of referents in memory more than their quantity, unlike capacity-based theories, which hypothesize that the number of referents play a key role in language processing.

1.3.3 Maintenance and the Sustained Anterior Negativity (SAN)

Partially due to the success of the LV05 model, the majority of studies examining WM processes in sentence comprehension have focused on the retrieval process (Dillon et al., 2013; Jäger et al., 2017; Van Dyke and McElree, 2006; Wagers et al., 2009), especially since many psycholinguistic phenomena explored occur at the retrieval site. WM encoding was mainly examined as a cause for interference at retrieval (encoding interference), as discussed in subsection 1.3.2 although a few works have looked at different processing costs at the encoding site (Hofmeister, 2011; Hofmeister and Vasishth, 2014). WM maintenance was also under-explored in the sentence comprehension literature using behavioral experiments although there were exceptions. A key question about the maintenance site that theorists are interested in is whether referents in WM are actively maintained or not (Ness and Meltzer-Asscher, 2017,1; Wagers and Phillips, 2014). Active maintenance, as discussed by Wagers and Phillips (2014), means that the language system spends extra processing resources to prevent these representations from decaying. Alternatively, if we look at this from the embedded-process model (Cowan, 1999), referents being actively maintained can also mean that they remain in WM instead of being displaced to activated LTM, which also can lead to more challenging retrieval. The current consensus in the literature is that only certain features of the referents, especially syntactic ones, are actively maintained (Ness and Meltzer-Asscher, 2017,1; Wagers and Phillips, 2014). The retrieval of semantic features of these referents was delayed compared to their syntactic counterparts when there was a lot of intervening material since encoding. The maintenance process and the specific question of active maintenance will be discussed further in subsections 2.1.1 and 3.1.1, respectively.

However, most of the interest in WM maintenance in sentence comprehension comes from the ERP literature, especially due to the presence of a persistent signal called the SAN (Fig 1.2). The SAN was first described by King and Kutas (1995) when they measured participants' EEG while reading sentences with subject RCs and object RCs. King and Kutas (1995) found that relative to reading subject RCs, reading object RCs elicited a sustained negativity that lasted from the point of divergence between conditions to the end of the sentence. This negativity was mainly observed in frontal electrodes.

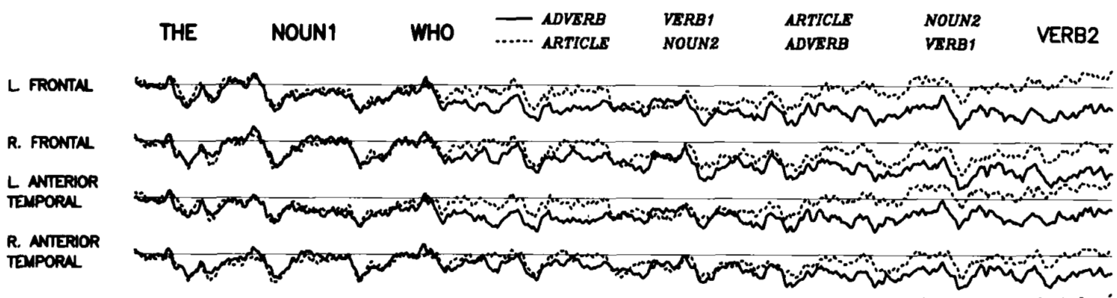

Figure 1.2: The sustained anterior negativity. SOURCE: King and Kutas (1995). Reproduced with permission from *MIT Press.*

Since this study, other experiments investigating more memory-intensive structures against less memory-intensive structures have reported the SAN (Cruz Heredia et al., 2022; Fiebach et al., 2002; Hagiwara et al., 2007; Kwon et al., 2013; Phillips et al., 2005; Ueno and Garnsey, 2008; Yang et al., 2010; Yano and Koizumi, 2018), although its onset and offset, its topographical distribution and whether its amplitude changes throughout the dependency were somewhat inconsistent across experiments. Some experiments examining structures with different demands to WM also did not detect the SAN (Lo and Brennan, 2021; Sprouse et al., 2021; Yano and Koizumi, 2021) or instead observed a more posterior sustained signal (Cruz Heredia et al., 2022). This lack of robustness has prompted many to question the functional interpretation of the SAN as a marker for WM maintenance load (Cruz Heredia et al., 2022; Lau and Liao, 2018; Sprouse et al., 2021). More discussion on the SAN and the debate over its functional interpretation can be found in subsection 3.1.2. Since this dissertation takes special interest in the maintenance process, an understanding of the SAN is crucial, especially for the two EEG experiments, Experiment 3 and Experiment 5. This dissertation aims to not only use the SAN to better understand WM maintenance during sentence comprehension but also contribute to the conversation on how to interpret the presence and absence of the SAN.

1.4 Representational Complexity, Internal Coherence and Similarity

To understand how linguistic representations in the context of sentence comprehension and their features are encoded, retrieved and most importantly, maintained in WM, all of my experiments manipulated properties that involve features of the referent, namely representational complexity and internal coherence. **Representational complexity**, in the current dissertation, is defined based on the definition in Hofmeister (2011). Representation A is more complex than representation B if the features making up representation B are a subset of those making up representation A. For example, *a nurse* is more representationally complex than *a person* because *a nurse* has all the features of *a person* and also the additional feature of being a nurse. While Hofmeister (2011) operationalized complexity based on both syntactic and semantic features, this dissertation will mainly consider semantic features. All experiments in this dissertation investigated the role of representational complexity in WM processes by making NPs more complex in two ways: pre-modification and coordination. A pre-modified NP *the hard-working lawyer* is more complex than a non-pre-modified NP *the lawyer* because the former has the additional feature of being hard-working. A coordinated NP *the lawyer and the judge* is more complex than its uncoordinated counterpart *the lawyer* because the former has the additional feature of being with the judge.

Previous works (Hofmeister, 2011; Hofmeister and Vasishth, 2014; Karimi et al., 2019,2; Karimi and Ferreira, 2016; Karimi et al., 2014,1; Troyer et al., 2016) have examined how NPs of different representational complexity have influenced the cost to encode them and to retrieve them. These studies have used adjectival/nominal pre-modification and relative clauses to make NPs more complex and have examined retrieval at either a gap site or during pronoun resolution. The prevailing pattern is that more complex NPs take more effort to be encoded but shorter to be retrieved. The studies did not discuss the influence of representational complexity on the maintenance site. There are two main accounts for the facilitation effect observed at retrieval (Hofmeister, 2011; Karimi et al., 2020). Firstly, it is possible that the longer encoding time of the representation makes it more salient in WM or gives it more attentional resources, leading to its higher activation level and more ease of retrieval (Karimi et al., 2020). Secondly, it is also possible that the higher number of features of the complex representation makes it more distinct from competing referents and less vulnerable to inhibitory interference effects. More on representational complexity and its effect on WM processes will be discussed in subsection 2.1.2.

Internal coherence, in the current dissertation, is defined as the compatibility between features within a representation. This compatibility can be the result of co-occurrence between the features based on the experience of the comprehender and/or world knowledge. For example, *an evil monster* is more coherent than *a kind monster* because monsters are often associated with being evil instead of kind. The experiments in this dissertation examining internal coherence, Experiments 4a, 4b and 5, made use of coordinated NPs made up of multiple referents. To make coordinated NPs with different internal coherence, the similarity between the referents and more specifically, whether the referents come from the same conceptual category, was manipulated. An example coherent NP is *the hamster, the cat and the dog* where the three referents come from the same conceptual category of animals, share more features between them and are more likely to occur together. An example incoherent NP is *the kayak, the TV and the dog* where the kayak comes from the category of transportation, the TV from the category of household object and the dog from the category of animal. Thus, they share fewer features among themselves and are less likely to occur together.

The role of internal coherence of representations on WM processes was under-investigated in sentence comprehension with the only exception of Experiment 3 from Hofmeister (2011). He observed that lower coherence of an NP can increase both encoding and retrieval effort, showing that having more complex NP is not always beneficial. On the other hand, as discussed in section 1.3, similarity is also studied extensively in the sentence comprehension literature through investigations of similarity-based interference. While that literature is informative for understanding internal coherence, the key difference is that the interference literature manipulates similarity between competing representations, not similarity between referents that are part of a representation. Further discussion about internal coherence and similarity in both contexts will be discussed in subsection 4.1.2.

1.5 A Note on Terminologies

Throughout the dissertation, for the sake of practicality and ease of reading, I will use the terms *encoding*, *maintenance* and *retrieval* to index very specific periods of time that will be defined in the corresponding experiment's Methods section. Roughly, the encoding period refers to the entirety of the target NP where the NP is brought into WM. The retrieval period refers to the verb that requires the retrieval of the target NP for interpretation and its subsequent spillover words. The maintenance period refers to the words in between the words of the encoding and retrieval periods. It is important to note that the demarcation between these three processes are not as clear. It is possible that parts of the target NP are already maintained in WM while other parts are still being encoded. Another scenario is where other referents are being encoded into WM during the maintenance period of the target NP. Yet another case is where retrieval happens preemptively during the maintenance period due to prediction. Thus, these three processes are dynamic and not confined to the boundaries set by my usage of terminologies. Nevertheless, I will point out and discuss whenever this concurrence between the processes is relevant for interpretation of the results.

1.6 Outline of the Dissertation

The dissertation investigates how the representational complexity and internal coherence of the target NP affects the effort associated with its encoding, maintenance and retrieval, with a particular focus on the maintenance process. In doing so, it aims to further elucidate the architecture of WM in sentence comprehension, especially with regards to how features of linguistic representations are processed in WM. It also seeks to draw insights from the

WM literature in other domains to better understand the structure of WM during sentence comprehension. This dissertation consists of 6 self-paced reading (SPR) and 2 electroencephalography (EEG) experiments. The examination of both behavioral and neural data gives us a multidimensional perspective into the three WM processes. The following four chapters include a more thorough discussion of the relevant background literature, the experimental procedure and stimuli, and presentation and discussion of the results for the 8 experiments, as well as a synthesis of the experimental results in the final chapter.

Chapter 2 presents 4 SPR experiments that investigated how the encoding, maintenance and retrieval of the target NP are modulated by its representational complexity. Specifically, Experiments 1a and 1b followed up on Hofmeister (2011) and used NPs pre-modified with adjectives or nouns as complex NPs. Experiments 2a and 2b instead used coordinated NPs as complex NPs to examine whether the effects observed in Experiments 1a and 1b were generalizable to other types of complex NPs. These experiments also manipulated whether the retrieval site was the verb of a subject or object RC to change the number of intervening referents introduced between encoding and retrieval. Overall, the experiments in this chapter showed inconsistent effects between using pre-modified NPs and coordinated NPs. Encoding of pre-modified NPs was slower than their simple counterparts but encoding of coordinated NPs was faster. The previously reported facilitatory effect at retrieval (Hofmeister, 2011; Karimi et al., 2020) was only observed for pre-modified NPs, not coordinated NPs. Most critically, the experiments showed a robust facilitatory effect during maintenance when the target NP is complex. They also demonstrated that this facilitation was the result of both increased salience/attentional resources for complex NPs and increased distinctiveness between the target NP and intervening referents.

Chapter 3 presents an EEG experiment that followed up on Experiments 1a and 1b from the neural perspective. Since many previous studies on WM maintenance in language comprehension were EEG studies, this experiment took advantage of the understanding of the SAN to study how complex pre-modified NPs are maintained in WM. It also explored whether there were ERP differences during encoding and retrieval of complex and simple NPs. The experiment showed that reading sentences with simple NP, relative to reading sentences with complex NP, showed a global negativity that lasted throughout the maintenance period. This result provided further evidence for the facilitatory effect of higher representational complexity during maintenance. Additionally, the experiment also showed an unexpected pattern of sustained differences between reading object RCs and subject RCs: Reading object RCs elicited sustained anterior positivity during maintenance and sustained anterior negativity during retrieval.

Chapter 4 presents 2 SPR experiments and an EEG experiment that investigated the encoding, maintenance and retrieval of complex NPs that varied in internal coherence. In the behavioral Experiments 4a and 4b and the EEG Experiment 5, internal coherence of the coordinated filler NP was manipulated by changing whether the coordinated referents came from the same conceptual category. Additionally, to observe the relationship between complexity and coherence, representational complexity was manipulated by modulating the number of coordinated referents. All three experiments showed that increasing the coherence within the target NP facilitates encoding and early maintenance of the NP in WM, while having limited effect during retrieval. These experiments also showed faster RT or less anterior negativity during maintenance when the target NP had more conjoined referents, further reinforcing the facilitatory effect during maintenance for more representationally complex NPs.

Chapter 5 finally summarizes the findings from the eight experiments and synthesized them into implications for the architecture of WM during sentence comprehension and also for functional interpretations of the SAN. The chapter also proposed limitations of current studies and future directions to further explore the intersection between language comprehension and WM through both behavioral and neural perspectives.

CHAPTER 2
THE EFFECT OF REPRESENTATIONAL COMPLEXITY ON WORKING MEMORY PROCESSES - A BEHAVIORAL PERSPECTIVE

2.1 Introduction

WM processes - encoding, maintenance and retrieval - are essential for sentence comprehension, especially for understanding long-distance dependencies, such as those present in relative clauses, pronoun resolution and other complex constructions. For example, consider the sentence below which contains an object-extracted relative clause:

(7) **The students** who Priyanka wholeheartedly **praised** submitted the report.

There are multiple syntactic dependencies in this sentence, one of which is the long-distance dependency between the noun phrase (NP) *the student* in the matrix clause and the embedded verb *praised*. In order to understand (7), comprehenders have to encode the matrix subject NP *the students* into WM, maintain it for a period of time until the other end of the dependency, i.e. the verb *praised*, appears in the sentence. At the verb *praised*, which is looking for an object, the matrix subject NP *the students* is retrieved and integrated into the object position.

There is a large body of previous work investigating the memory retrieval mechanism during online sentence processing (Dillon et al., 2013; Jäger et al., 2017; Lewis and Vasishth, 2005; Van Dyke and McElree, 2006; Wagers et al., 2009). Many of these studies investigated the extent to which the syntactic or semantic similarity between the different referents in the sentence can have on retrieval cost. When the target referent is syntactically or/and semantically similar to the competing referents, retrieval can be more costly.

Comparatively speaking, the mechanisms of WM encoding and maintenance in sentence processing are less explored. For the encoding process, there is some evidence that interference during WM encoding contributes to increased retrieval cost (Barker et al., 2001; Gordon et al., 2001; Hofmeister and Vasishth, 2014; Kush et al., 2015; Villata et al., 2018). Many theorists argued that when referents share similar features, encoding these referents into WM will degrade their representations, leading to lower-quality representations and higher effort to retrieve them (Nairne, 2002; Oberauer and Kliegl, 2006). However, not many studies have examined the cost associated with the encoding process itself (with the exceptions of Hofmeister (2011) and Hofmeister and Vasishth (2014), which will be described in section 2.1.2). For the maintenance process, most of the understanding came from EEG studies, as well as a few behavioral studies, which will be described in the following subsection.

2.1.1 Working Memory Maintenance in Sentence Comprehension

One influential idea on WM maintenance is that since WM is capacity-limited (Cowan, 2010; McElree, 2001; Oberauer, 2002), the more items that need to be stored in WM, the higher the maintenance cost (Cowan, 2010). The magnitude of many sustained EEG signals in the visual and auditory domains are modulated by WM load - the number of items in the memorandum (Lefebvre et al., 2013; Nolden et al., 2013; Vogel and Machizawa, 2004).

In the domain of sentence comprehension, major theories about processing cost that engage with WM processes either did not refute or did not acknowledge the existence of a storage cost during maintenance. Additionally, these theories usually make predictions about the cost incurred during WM retrieval, not during WM maintenance itself. The capacity-based models advocated by Gibson (1998, 2000) consider two types of cost that modulate retrieval difficulty: storage cost and integration cost. However, storage cost under these models is not regarding the storage of existing representations or referents in WM but of predicted syntactic heads in order to fulfil incomplete dependencies. A phenomenon that these models attempt to explain is the reading time slowdown on the RC verb *criticized* in (9) relative to (8) (the subject RC advantage). At the verb in (8), there are two predicted heads to fulfil the dependency: the RC object noun and the RC verb; however, in (9), there are three: the RC subject noun, the RC verb and the matrix verb. Thus, more storage cost is needed in (9), leading to the increased retrieval cost. On the other hand, integration cost, the cost of integrating the incoming word into the syntactic structure, is modulated by the number of intervening referents between the filler and the gap. In (9), the referent *the witness* intervenes between *who* and *criticized*, leading to more decay of the filler in WM relative to in (8), where there is no intervening referent. While this is more similar to the idea of the number of items in WM, its nature as integration cost cannot inform us about the cost during maintenance.

(8) The lawyer who **criticized** the witness admitted the error.

(9) The lawyer who the witness **criticized** admitted the error.

The other major theory, the cue-based retrieval framework (Lewis and Vasishth, 2005), while again focusing on effects during WM retrieval, denies the existence of storage-based maintenance cost based on the results of Van Dyke and Lewis (2003). Van Dyke and Lewis (2003) considered pairs of non-garden-path sentences such as (10) and garden-path sentences such as (11). In (10), when participants read the RC *who said the townspeople are dangerous*, they predict an upcoming verb for the *that*-clause. In (11), participants were garden-pathed into interpreting *the man* as the object of the matrix verb *understood* and thus, did not predict an upcoming verb. If there was storage cost during maintenance, RT during the RC would be predicted to be slower when participants read (10) than when they read (11) since they had to maintain an extra prediction of the verb and thus incur higher storage cost in (10). However, Van Dyke and Lewis (2003) found no significant differences in RT between the two conditions for the RC *who said the townspeople are dangerous*. This result made Lewis and Vasishth (2005) doubt the presence of storage cost during WM maintenance.

(10) The frightened boy understood that the man who said the townspeople are dangerous was paranoid about dying.

(11) The frightened boy understood the man who said the townspeople are dangerous was paranoid about dying.

Nevertheless, there has been a handful of studies that suggested the existence of maintenance cost. This is particularly the case in the EEG literature, especially with respect to a component called the sustained anterior negativity (SAN). The SAN is a negative slow wave that persists throughout the maintenance period. Previous experiments comparing between structures with different WM demands (Fiebach et al., 2002; King and Kutas, 1995; Phillips et al., 2005) showed presence of the SAN in the more memory-intensive structure relative to the less memory-intensive structure. For instance, King and Kutas (1995) found the SAN when comparing the EEG elicited when participants were reading a sentence with an ORC (13) to when they were reading a sentence with an SRC (12). The SAN they observed lasted from the first word of the RC (*harshly* in (12) and *the* in (13)) to the end of the sentence.

(12) The reporter who harshly attacked the senator admitted the error.

(13) The reporter who the senator harshly attacked admitted the error.

Other examples of memory-intensive structures that have been reported to show the SAN are German object *wh*-phrases relative to subject *wh*-phrases (Fiebach et al., 2002), English *wh*-phrases relative to *that*-phrases (Phillips et al., 2005), Japanese scrambled word order relative to non-scrambled order (Hagiwara et al., 2007; Yano and Koizumi, 2021) and temporal clauses with *before* relative to those with *after* (Politzer-Ahles et al., 2017).

However, some recent works have questioned this interpretation of the SAN and suggested alternative interpretations of the SAN (Lau, 2018; Yano and Koizumi, 2021). Yano and Koizumi (2021) found that when comparing between Japanese scrambled and non-scrambled word order, the SAN can be eliminated if the non-canonical scrambled word order is licensed by context, which poses explanatory difficulty for the WM account. Lau (2018) pointed out that the details of the SAN found in previous studies are not consistent with each other and with the fMRI activity in the left inferior frontal cortex, which has been hypothesized to correlate with the SAN behavior (Fiebach et al., 2005).

In terms of behavioral measures, a recent eye-tracking study by Ristic et al. (2021) in both Spanish and English reported evidence consistent with maintenance cost. They compared various eye-tracking metrics during the embedded clause as participants read sentences such as (14), where participants have to store *that student* in anticipation of the verb, and (15), where they do not have to store *that student*. In English, the study found marginal effect for first-pass reading time and significant effect for go-past reading time over the entire embedded clause where both metrics are longer in (14) than in (15).

(14) That student, *as soon as a professor finishes the class,* leaves the classroom.

(15) I watched that student, *and as soon as a professor finishes the class,* she leaves the classroom.

Collectively, these studies showed that maintaining a filler in a filler-gap dependency in WM results in an increase in processing cost, typified by either a sustained ERP signal or increased RTs. These studies all manipulated syntactic complexity by making significant changes to the sentence structure. However, in the following experiments, we instead focused on a smaller change to the complexity of a single linguistic representation, not the whole sentence and observed how that change affected the cost of WM processes, especially maintenance.

2.1.2 Representational Complexity and Working Memory

There are different ways to define the complexity of a representation. For the current purpose, a linguistic representation is considered to be more complex than another if it has more syntactic and/or semantic features. For example, an NP modified by an adjective such as *a Venezuelan communist* is representationally more complex than its counterpart that is not modified by an adjective such as *a communist.* Not only is the noun phrase *a Venezuelan communist* syntactically more complex than the simpler noun phrase *a communist*, the former is also semantically/conceptually richer and hence expresses a more specific concept than simply a communist. A more complex representation does not have to be both syntactically and semantically more complex. The NP *a communist*, while having a similar syntactic structure, is semantically richer than the NP *a person.*

A number of previous studies have examined how varying the representational complexity of an NP referent can modulate processing cost, specifically, during retrieval from WM and to a limited extent, encoding into WM (Hofmeister, 2011; Hofmeister and Vasishth, 2014; Karimi et al., 2019,2; Karimi and Ferreira, 2016; Karimi et al., 2018; Troyer et al., 2016). The first study in this line of research, Hofmeister (2011), compared NPs that are pre-modified by nouns or adjectives (*an alleged Venezuelan communist* with NPs that are not (*a communist*). The following pair of examples were adapted from Hofmeister (2011).

(16) It was **a communist** who the members of the club <u>banned</u> from ever entering the premises.

(17) It was **an alleged Venezuelan communist** who the members of the club <u>banned</u> from ever entering the premises.

For sentences like (16) and (17) with a simple and complex matrix subject NP, respectively, Hofmeister (2011) showed that reading time (RT) on the word *communist* is slower in (17) than in (16). But at the same time it was also found that on the word right after the verb *banned*, the RTs were faster in (17) than (16). This was taken as evidence that encoding a complex NP is more costly than encoding a simple one, but a more effortful encoding can facilitate memory retrieval later at the verb when the NP referent needs to be retrieved and integrated (Hofmeister, 2011; Hofmeister and Vasishth, 2014).

Troyer et al. (2016) showed that a similar facilitatory effect can be observed across sentences. In this experiment, participants read paragraphs consisting of several sentences. The penultimate sentence (18) described two target referents in different amounts of details. In this example, the senator the Republican had voted for was described in greater details (*a man from Ohio who was running for president*) than the senator the Democrat had voted for. The final sentence (19) required the reader to retrieve one of the two referents. Troyer et al. (2016) found that the RT for the retrieval site *had voted for* was faster if the referent was more elaborated in (18) (*the senator who the Republican had voted for*). They concluded that greater elaboration of a referent during discourse can also facilitate its retrieval.

(18) The Democrat had voted for one of the senators, and the Republican had voted for the other, a man from Ohio who was running for president.

(19) The senator who the {Republican/Democrat} had voted for was picking a fight about health care reform.

Karimi and Ferreira (2016) showed, through two eye-tracking experiments using the visual world paradigm, that when confronted with an ambiguous pronoun, participants tend to look at the more representationally complex referent (*the wizard who was confused and depressed by the irreparable situation*) that is modified by a relative clause in the previous sentence than the unmodified referent (*the knight*). In language production, participants also are more likely to use a pronoun to refer to the more representationally complex referent than its simple counterpart (Karimi et al., 2014). Both of these results pointed to the higher accessibility of more complex representations compared to their simpler counterparts, which complemented findings of facilitated retrieval of complex representations in SPR experiments (Hofmeister, 2011; Hofmeister and Vasishth, 2014; Troyer et al., 2016).

These experimental results together demonstrated higher processing cost of complex representations during encoding but lower processing cost during retrieval. The higher encoding cost can be attributed to the effort associated with encoding more syntactic structures and/or more semantic features into WM. Regarding the lower retrieval cost, there are various possible explanations. Karimi and colleagues argued that less effort is needed to retrieve a complex representation because it is more accessible the simple counterpart. Using the cue-based retrieval framework (Lewis and Vasishth, 2005), Karimi et al. (2014) and Karimi and Ferreira (2016) hypothesized that repeated activations of the more complex referent during elaboration increases its activation level and makes the more complex referent more accessible. Additionally, the more complex representation also contains more features and can satisfy more retrieval cues, making it more accessible (Craik and Tulving, 1975). Another possibility, discussed in Karimi et al. (2020), is that because more time is spent on encoding complex NPs, which are also longer, more attentional resources are dedicated to complex NP referents, leading to higher activation level and consequently, less effortful retrieval.

An alternative but not mutually exclusive explanation for the decreased retrieval cost is that more complex representations, which have more features, are more distinct from competing referents and thus, are less subjected to similarity-based retrieval interference. Previous literature has shown that retrieval can be more difficult if the target referent and other referents in WM are syntactically and/or semantically similar (Cunnings and Sturt, 2014,1; Mertzen et al., 2020; Nicenboim et al., 2018). When a representationally complex referent has an additional feature that is not shared with competing referents, it is more dissimilar from its competitors than its simple counterpart, rendering its retrieval less costly.

While aforementioned works have addressed many aspects regarding how varying representational complexity can modulate retrieval costs, there are many questions remained unanswered or under-explored. Since the bulk of the works in this domain concerns the retrieval process and to a less extent, the encoding process, little is known about how representational complexity affects the third WM process, maintenance. In addition, with the exception of Karimi et al. (2020), little attention has been paid to disentangling the different possible explanations outlined in the previous paragraphs. Lastly, previous studies have largely focused on representational complexity by means of modification, either using adjectives or relative clauses. Thus, it is unclear whether the observed effects can be generalized as directly related to representational complexity.

Our study aimed to address the questions outlined in the previous paragraph. In four experiments, we sought to 1) conceptually replicate the reported effects of target NP complexity during WM encoding and retrieval and 2) explore the cost of maintaining representationally complex NPs. To observe WM maintenance of target NPs, the experiments employed a maintenance window between the encoding and retrieval sites. We hypothesized that similar to retrieving complex representations, maintaining complex representations is also less costly. Additionally, we manipulated the number of competing referents to the target referent to examine whether increasing representational complexity facilitates WM processes by increasing distinctiveness between the target referent and its competitors. By looking at RT differences in the different time windows, we would be able to study the effort associated with encoding, maintaining and retrieving complex representations and understand which mechanisms account for differences in processing effort. Lastly, while the first two experiments, 1a and 1b, made use of modification to increase complexity, the last two experiments, 2a and 2b, investigated coordinated NPs to examine which effects are generalizable and which effects are specific to modified NPs.

2.2 Experiment 1a

2.2.1 Participants

101 participants aged 19-50 (mean age = 30.8), recruited on Prolific, participated in this experiment for payment. All participants self-identified as native, monolingual English speakers who were raised in monolingual households. In addition, participants had no language related disorders or literacy difficulties. Data from 17 out of 101 participants was removed because they scored below the threshold of 75% for the comprehension questions (1 standard deviation below the mean accuracy, rounded to the nearest multiples of 5%).

2.2.2 Methods and Materials

The self-paced reading experiment had a 2 x 2 design. It consisted of 32 4-condition items which were manipulated in terms of the complexity of the matrix subject filler NP (Simple vs. Complex) and the type of relative clause (RC) involved (subject-extracted (SRC) vs. object-extracted (ORC)). An example is given in (20a)-(20d), with the slashes indicating the self-paced reading regions. In the complex NP condition, the filler NP contains two modifiers, whereas in the simple NP condition, the filler NP does not contain any modifier.

All sentences consisted of the matrix subject filler NP, which was followed by the RC modifying the filler NP, then an adverb of time, the matrix verb and the matrix object NP. The filler NP always started with *those* and was modified by either 0 (simple) or 2 (complex) nouns or adjectives. In both types of RC manipulations, the RC verb was preceded by an adverb to allow for an extended maintenance window. Example sentences for Experiment 1a are as follows:

(20) a. **Complex, SRC:** Those / emotional / crash / survivors / who / dutifully / assisted / Sophia / last week / joined / the meeting.

b. **Complex, ORC:** Those / emotional / crash / survivors / who / Sophia / dutifully / assisted / last week / joined / the meeting.

c. **Simple, SRC:** Those / survivors / who / dutifully / assisted / Sophia / last week / joined / the meeting.

d. **Simple, ORC:** Those / survivors / who / Sophia / dutifully / assisted / last week / joined / the meeting.

Each participant read only one condition per item, totaling 32 experimental sentences. In addition, they also read 32 filler sentences, which were sentences of various types that contained long distance dependencies. Each sentence, experimental or filler, preceded a yes-no comprehension question targeting the dependency between the RC verb *assisted* and either the filler NP, e.g. *those emotional crash survivors/those survivors* or the NP introduced in the RC, e.g. *Sophia*. An example question for the sentence set above is *Was it those survivors who assisted Sophia?*. The expected answer for half of the questions was *Yes* and for the other half was *No*. Participants did not receive feedback on the accuracy of their answers. In this experiment, average accuracy across all items (including fillers) was 85.9%, and the average accuracy on the experimental sentences was 84.0%.

The experiment was carried out on Ibex Farm (Drummond, 2013). Participants did two practice trials before reading the experimental and filler items in a randomized order. Before a trial started, a dash line appeared in the middle of the screen where the stimuli would appear. Upon pressing the space bar, the dash line disappeared and the first word appeared. Participants were instructed to press the space bar to continue reading the sentence. As the space bar was pressed, the current word(s) was replaced by the subsequent word(s).

For statistical analysis, after excluding data from participants who did not meet the comprehension accuracy threshold of 75%, raw RT beyond three standard deviations of the mean raw RT at each sentence position and condition were excluded. We rejected 1.03% of the raw RTs through this procedure. Following Hofmeister and Vasishth (2014), we did not exclude RTs of sentences whose comprehension question was answered incorrectly. This was done to ensure we did not discard instances that might have reflected failure to maintain the correct dependency in WM.

RTs in each region were log-transformed and then residualized on two predictors: the linear position of a region in a sentence and log RT of the region immediately prior to the current one. Both predictors are known to impact self-paced reading RTs for independent reasons.

Residualization was done using linear mixed models through the *lmer* package (Bates et al., 2015). RTs from fillers were included in the residualization process. Residualized log RTs served as the dependent variable for our analyses.

For data analysis, we considered three sites of interest - encoding, maintenance and retrieval sites. The **encoding site** included the head noun of the filler NP, i.e. *survivors*. RTs at this region reflected processing effort to encode the subject noun phrase into WM representation. The **retrieval site** included the RC verb *assisted* and the spill-over region after the RC verb, i.e. the temporal adverb *last week*. At the retrieval site, the filler NP needs to be retrieved from WM to serve as the subject or object of the RC verb. We also analyzed the **maintenance site**, which included the words between the encoding and retrieval sites: the complementizer *who*, the RC subject NP *Sophia* in ORCs, and the preverbal adverb *dutifully*. During this period, the filler NP is maintained in WM, awaiting for retrieval.

For statistical analyses, we employed Bayesian hierarchical modeling, using the R package *brms* (Bürkner, 2017). For each self-paced reading region examined, the model used 4 chains, with 2000 samples per chain, the initial 1000 samples being warm-up samples and no thinning. This led to 4000 post-warmup samples for each parameter estimate per region. Models of regions up to (and including) the complementizer *who* comprised of the fixed effect of filler NP complexity (sum coded, simple -0.5, complex 0.5). Models of regions after the complementizer *who* comprised of fixed effects of filler NP complexity (sum coded, simple -0.5, complex 0.5), RC type (sum coded, SRC -0.5, ORC 0.5) and their interaction. All models also contained by-participant and by-item random intercept adjustments and random slopes for all fixed effects analyzed in that region. We used relatively weak, uninformative priors for all parameters. For the prior for all the fixed effects, including the intercept, we used a normal distribution N(0,10) with mean 0 and standard deviation of 10[1]. The final results

1. An example of a *brm* model:
brm(formula = residualizedlogrt ~ npcomplexity * rctype + (1 + npcomplexity * rctype|participant) + (1 + npcomplexity * rctype|item), data = rt_data, family = gaussian(),
prior = c(prior('normal(0,10)', class = 'Intercept'), set_prior('normal(0,10)', class = 'sigma'), set_prior('normal(0,10)', class = 'b'),set_prior('normal(0,10)', class = 'sd'), set_prior('lkj(2)', class = 'cor')),
warmup = 1000, iter = 2000, chains = 4, control = list(adapt_delta = 0.99, max_treedepth = 12))

were reported in terms of the mean of the posterior distributions and the 95% credible intervals. We considered a predictor as reliable if the credible interval did not include 0.

2.2.3 Results

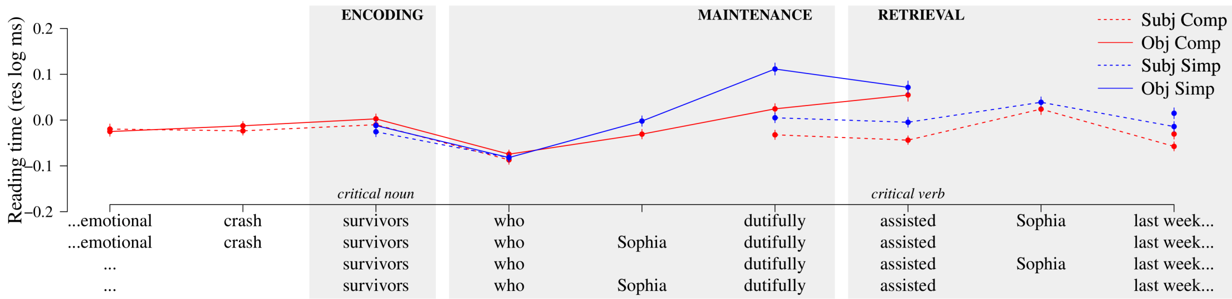

Figure 2.1: Experiment 1a: Residualized log reading times

In the **encoding site**, as shown in Figure 2.2 and Table 2.1, there was no effect of NP complexity on the head noun *survivors*, contrary to the slowdown for more complex NPs observed by Hofmeister (2011) and Hofmeister and Vasishth (2014).

In the **maintenance site**, there was no effect of NP complexity on the complementizer *who*. On the preverbal adverb *dutifully*, there was evidence for effects of filler NP complexity and RC type, as well as weak evidence for an interaction. More specifically, participants read this word faster if the filler NP is complex. They also read faster if they were reading an SRC rather than an ORC. The weak evidence for an interaction arose from the fact that the facilitation effect due to NP complexity is more pronounced in ORCs than SRCs.

In the **retrieval site**, on the RC verb *assisted*, there was only evidence for the fixed effect of RC type. Similar to the maintenance site, RTs were faster in an SRC than in an ORC, which is an expected effect given previous works on the SRC advantage in English (Gibson, 1998, 2000; King and Just, 1991; Staub, 2010; Traxler et al., 2002). There continued to be evidence for the effect of RC type on the adverb *last week*. More critically, we only observed weak evidence of an effect of filler NP complexity on the RC verb *assisted*, with a faster RT in conditions with complex filler NP. However, this speed up was reliable on the spillover region, the adverb *last week*.

Region	Effect	Mean	CrI lower	CrI upper
survivors	Filler NP complexity	0.015	-0.021	0.048
who	Filler NP complexity	0.002	-0.023	0.027
dutifully	Filler NP complexity	-0.062	-0.087	-0.036
	RC type	0.079	0.043	0.114
	Filler NP comp x RC type	-0.045	-0.092	0.002
assisted	Filler NP complexity	-0.027	-0.058	0.003
	RC type	0.087	0.053	0.122
	Filler NP comp x RC type	0.024	-0.036	0.085
last week	Filler NP complexity	-0.044	-0.070	-0.020
	RC type	0.030	0.005	0.057
	Filler NP comp x RC type	-0.002	-0.052	0.047

Table 2.1: Experiment 1a: Model summary for each region and fixed effect or interaction examined

2.2.4 Discussion

We did not replicate the slowdown in the encoding site due to filler NP complexity, as previously observed (Hofmeister, 2011; Hofmeister and Vasishth, 2014). However, in the retrieval site, consistent with previous work (Hofmeister, 2011; Hofmeister and Vasishth, 2014; Karimi et al., 2020; Karimi and Ferreira, 2016), there was evidence in our results showing that having a complex filler NP speeds up RT.

More importantly, a novel finding from the current work is that having a more complex filler NP also benefited the maintenance process, as evident by the facilitation effect observed on the preverbal adverb *dutifully*. There was also some evidence that this effect was more pronounced in ORCs than in SRCs. The difference between the two clause types at the preverbal position may potentially speak to the mechanism based on which representational complexity can facilitate maintenance. In particular, when there are multiple competing referents that need to be maintained in WM, having a rich set of features on one referent helps keeping it apart from other referents. In SRCs, prior to the RC verb, there was only one referent, denoted by the filler NP *survivors*, maintained in WM. In ORCs, however, there were two competing referents in WM prior to the RC verb, as denoted by the filler NP *survivors* and the RC subject *Sophia*. Therefore, in ORCs, making one NP semantically rich (i.e. *the emotional crash survivors*) helped maintain the distinctions between the two

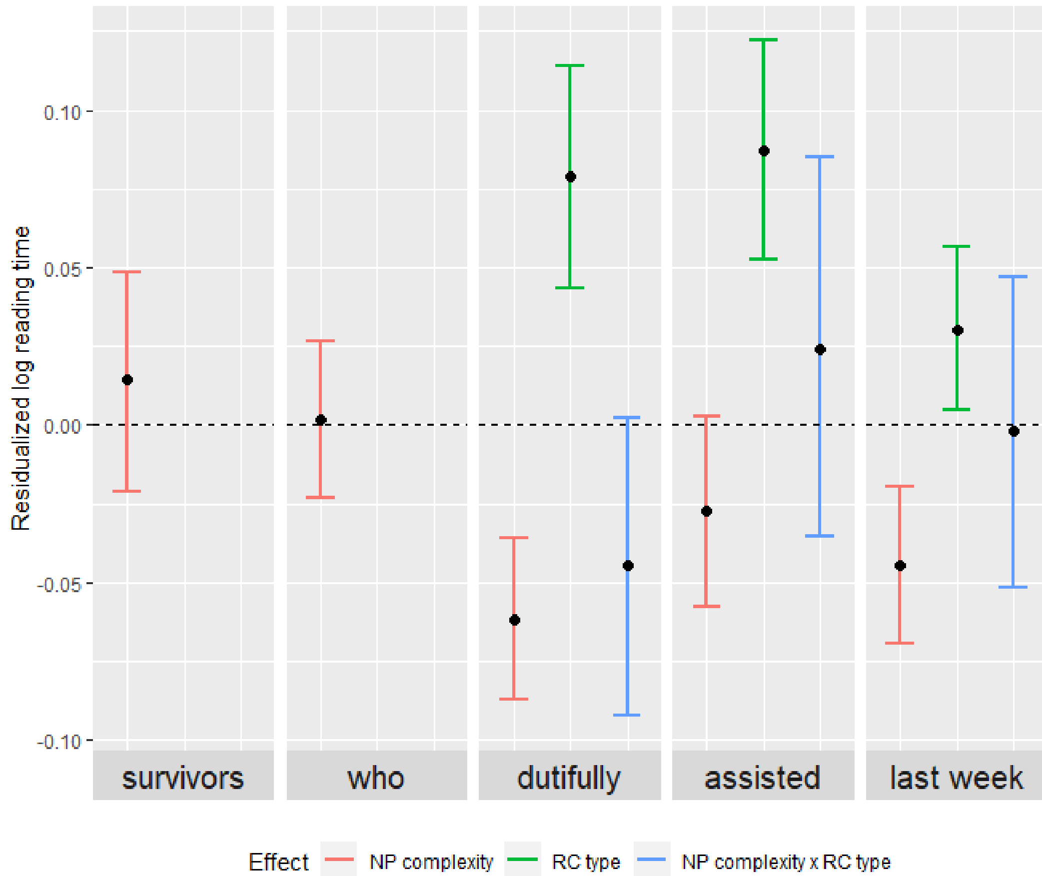

Figure 2.2: Experiment 1a: Credible intervals

referents. In SRCs, since there were no competing referents, whether the only referent maintained in WM is simple or complex did not make as big a difference as for the ORCs. But it is important to note that, even for SRCs, there was still a small effect, i.e., having a complex filler NP still resulted in faster RT on the word *dutifully* (β = -0.039, 95% CrI [-0.043, -0.008]). This suggests that a complex filler NP can benefit maintenance in ways that are independent from enhancing distinctiveness among competing representations in WM.

2.3 Experiment 1b

To test the hypothesis that featural richness helps maintain distinctiveness between referents, in Experiment 1b, we added an additional level of embedding before the RC verb to all experimental sentences. This was done so that by the preverbal adverb *dutifully*, for both SRCs and ORCs, there were multiple competing referents that need to be maintained in WM. We expected to see a larger facilitation effect for SRCs during the maintenance period due to filler NP complexity, potentially reducing (or even eliminating) the differences between SRCs and ORCs in terms of the facilitation effect.

2.3.1 Participants

100 participants aged 18-49 (mean age = 32.5), recruited through the data collection website Prolific, participated in this experiment for payment. All participants self-identified as native, monolingual English speakers who were raised in monolingual households. In addition, participants had no language related disorders or literacy difficulties. Data from 17 out of 100 participants was removed because they scored below the threshold of 65% (1 standard deviation below the mean accuracy, rounded to the nearest multiples of 5%) for the comprehension questions.

2.3.2 Methods and Materials

The experiment had the same setup and materials as Experiment 1a, except for the fact that an additional level of embedding *Jennifer thinks* was added after the complementizer *who* in all four conditions. Thus, Experiment 1b had the same 2 (NP complexity, sum coded, simple -0.5, complex 0.5) x 2 (RC type, sum coded, SRC -0.5, ORC 0.5) design. The additional NP was always a proper name. Example sentences for Experiment 1b are as follows:

(21) a. **Complex, SRC:** Those / emotional / crash / survivors / who / Jennifer / thinks / dutifully / assisted / Sophia / last week / joined / the meeting.

b. **Complex, ORC:** Those / emotional / crash / survivors / who / Jennifer / thinks / Sophia / dutifully / assisted / last week / joined / the meeting.

c. **Simple, SRC:** Those / survivors / who / Jennifer / thinks / dutifully / assisted / Sophia / last week / joined / the meeting.

d. **Simple, ORC:** Those / survivors / who / Jennifer / thinks / Sophia / dutifully / assisted / last week / joined / the meeting.

The additional level of embedding was also added to some fillers and all comprehension questions. An example comprehension question is *Was it those survivors who Jennifer thinks were assisted by Sophia?*. The sentences were presented in a self-paced reading task, like in Experiment 1a. In this experiment, average accuracy from all participants for the comprehension questions, including those following filler sentences, was 77.9%, and the average accuracy on the experimental sentences was 73.4%.

All statistical analyses followed the same procedure as Experiment 1a. We removed 1.24% of RT data that were 3 standard deviations of the mean raw RT at each sentence position and condition. The encoding and retrieval sites were the same as Experiment 1a. The maintenance site was extended to include the additional level of embedding *Jennifer thinks*.

2.3.3 Results

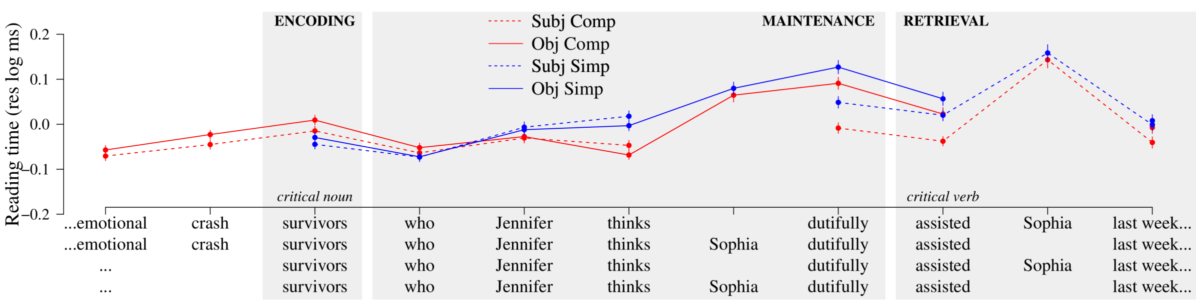

Figure 2.3: Experiment 1b: Residualized log reading times

In the **encoding site**, as demonstrated by Figure 2.4 and Table 2.2, on the filler head noun *survivors*, a main effect of NP complexity was observed. More specifically, RTs were slower when the filler NP was complex. This result was in line with the slowdown found by Hofmeister (2011) and Hofmeister and Vasishth (2014) but was different from Experiment 1a.

In the **maintenance site**, there was no main effect of NP complexity on the complementizer *who* and the noun *Jennifer*. However, on the following verb *thinks*, there was a main effect of NP complexity, where participants read faster if the head noun was complex. It is noteworthy because this verb is the first word by which two competing referents have been introduced into in WM, *survivors* and *Jennifer*.

Region	**Effect**	**Mean**	**CrI lower**	**CrI upper**
survivors	Filler NP complexity	0.034	0.008	0.062
who	Filler NP complexity	0.015	-0.011	0.041
Jennifer	Filler NP complexity	-0.020	-0.045	0.006
thinks	Filler NP complexity	-0.065	-0.092	-0.039
dutifully	Filler NP complexity	-0.047	-0.076	-0.018
	RC type	0.090	0.057	0.123
	Filler NP comp x RC type	0.020	-0.038	0.081
	Filler NP complexity (SRC)	-0.057	-0.097	-0.016
	Filler NP complexity (ORC)	-0.037	-0.079	0.006
assisted	Filler NP complexity	-0.044	-0.072	-0.017
	RC type	0.051	0.014	0.088
	Filler NP comp x RC type	0.027	-0.031	0.084
last week	Filler NP complexity	-0.027	-0.054	0.0003
	RC type	0.021	-0.016	0.056
	Filler NP comp x RC type	0.025	-0.031	0.082

Table 2.2: Experiment 1b: Model summary for each region and fixed effect or interaction examined

On the adverb *dutifully*, we found a main effect of NP complexity, where RTs were faster in sentences with a complex filler NP. We also found a main effect of RC type, where sentences with an SRC were read faster in this region. Unlike Experiment 1a, no evidence for an interaction between NP complexity and RC type was found. By the time this word appeared in the sentence, there were still two competing NP referents in SRCs (*survivors* and *Jennifer*) while there were three competing NP referents in ORCs (*survivors*, *Jennifer* and *Sophia*).

At the **retrieval site**, on the RC verb *assisted*, different from Experiment 1a, we found a main effect of NP complexity, where RTs were faster in sentences with a complex filler NP. This was consistent with previous results that complexity facilitates retrieval Hofmeister (2011) and Hofmeister and Vasishth (2014). There was also some weak evidence that this facilitation effect was extended to the spillover region, on the adverb *last week*. In additions, there was a RC type effect on the RC verb *assisted*, where sentences with SRCs were read more quickly. The main effect of RC type did not persist to the adverb *last week*.

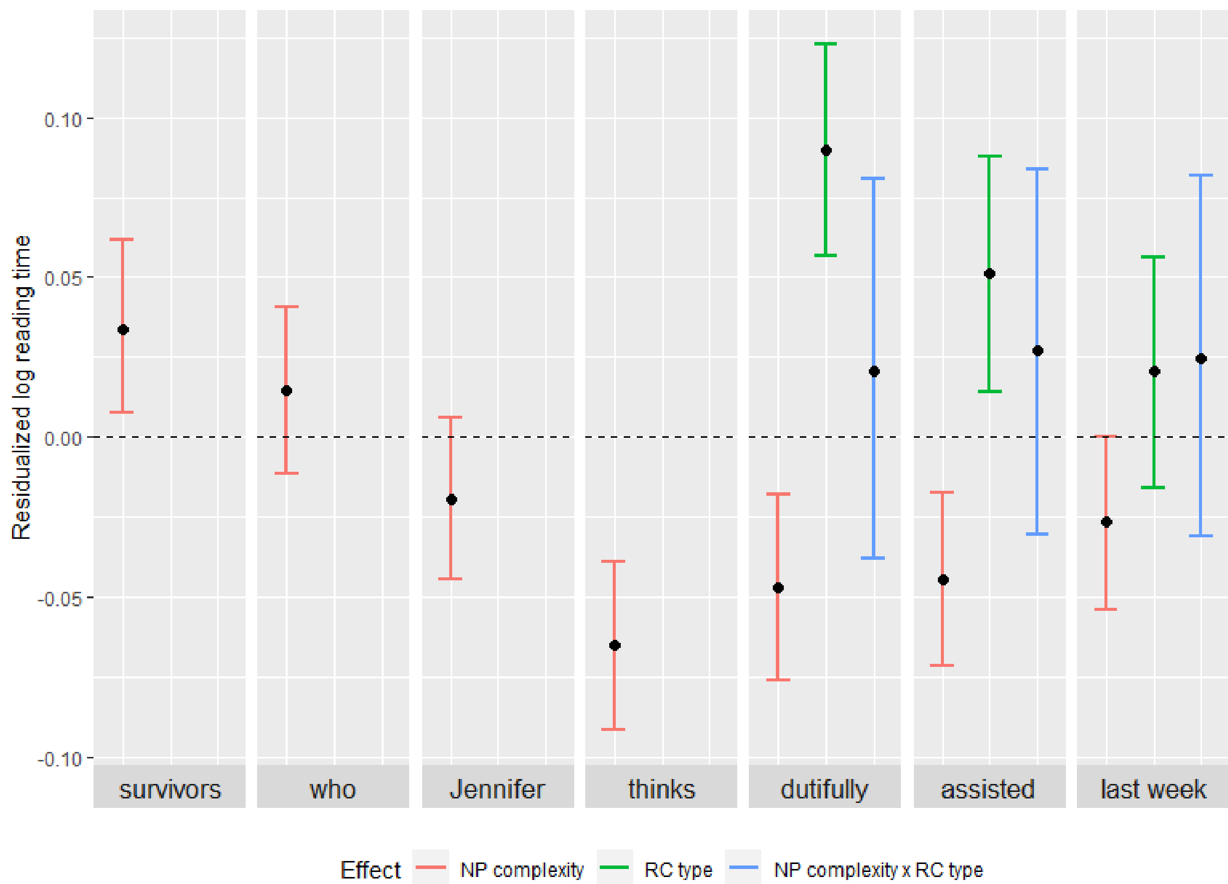

Figure 2.4: Experiment 1b: Credible intervals

2.3.4 Discussion

Different from Experiment 1a and consistent with results from Hofmeister (2011), we found a slowdown due to filler NP complexity in the encoding site. It is unclear why there was a difference between the two experiments since the sentence material up to the noun *survivors* was the same between the two experiments. This difference suggests that the increased effort during encoding might be less robust and might differ between subject pools. On the other hand, at the retrieval site on the RC verb *assisted* and to some extent on the spillover region *last week*, both experiments found the facilitation effect due to filler NP complexity.

Critically, Experiment 1b confirmed the hypothesis that a complex filler NP facilitates WM maintenance by enhancing the distinctiveness between representations of competing referents. In all conditions, by the time the verb *thinks* appeared in the sentence, there were two competing referents that need to be maintained in WM: *survivors* and *Jennifer*. For both SRC and ORC, the condition with a complex filler NP was read faster on the verb *thinks* than the condition with a simpler filler NP. Furthermore, the main effect of NP complexity continued on the adverb *dutifully*, and the interaction between filler NP complexity and RC type found in Experiment 1a disappeared. These results lend support to our hypothesis: since there are multiple competing referents during the maintenance period for both SRCs and ORCs, the benefit of having distinct representations could be observed in both types of clauses.

Experiments 1a and 1b demonstrated that having a complex pre-modified target NP increases encoding effort while decreasing maintenance and retrieval effort, which is consistent with findings from Hofmeister (2011). However, it is possible that these observed effects are specific to pre-modified NPs and not generalizable to other types of complex NPs.

In the next two experiments, Experiments 2a and 2b, we adapted the same design as Experiments 1a and 1b, respectively, to compare more complex coordinated NPs such as *those lawyers and judges* and their simpler counterparts such as *those judges*. By comparing and contrasting the results between these experiments, we can conclude which observed effects are the result of increasing representational complexity in general and not just pre-modification.

2.4 Experiment 2a

2.4.1 Participants

101 participants aged 18-50 (mean age = 30.8), recruited through the data collection website Prolific, participated in this experiment for payment. All participants self-identified as native, monolingual English speakers who were raised in monolingual households. In addi-

tion, participants had no language related disorders or literacy difficulties. Data from 14 out of 100 participants was removed because they scored below the threshold of 70% (1 standard deviation below the mean accuracy, rounded to the nearest multiples of 5%) for the comprehension questions.

2.4.2 Methods and Materials

Experiment 2a had a 2 (whether the filler NP is a coordinated NP or not) x 2 (SRC or ORC) design. All sentences started with *It seems that* or *It appears that* so that the filler NP is not placed at the beginning of the sentence. The simple uncoordinated filler NPs, took the form of *those plural-N*. The complex coordinated filler NPs took the form of *those plural-N and plural-N*. All the nouns in filler NPs in this experiment are human nouns. The remaining part of the sentences from the complementizer *who* onward had the same SRC/ORC structure as sentences in Experiment 1a. Example sentences are as follows:

(22) a. **Complex, SRC:** It seems / that / those judges / and lawyers / who / harshly / reprimanded / Andy / today / admitted / the error.

b. **Complex, ORC:** It seems / that / those judges / and lawyers / who / Andy / harshly / reprimanded / today / admitted / the error.

c. **Simple, SRC:** It seems / that / those lawyers / who / harshly / reprimanded / Andy / today / admitted / the error.

d. **Simple, ORC:** It seems / that / those lawyers / who / Andy / harshly / reprimanded / today / admitted / the error.

This experiment had 32 experimental sentences and 32 filler sentences, all of which were followed by a comprehension question. The filler sentences were the same as those in Experiment 1a. The comprehension questions for this experiment also had the same structure as those in Experiment 1a, in the form of *Was it those judges and lawyers who were reprimanded by Andy?*. Half of the questions were in active voice and half of the questions were in passive voice. The sentences were presented in a self-paced reading task in the same manner as in the previous experiments.

All statistical analyses followed the same procedure as Experiment 1a. We removed 0.87% of reading time data that were 3 standard deviations of the mean raw RT at each sentence position and condition. The encoding site in this experiment was defined as the last self-paced-reading chunk in the filler NP, i.e. the second conjunct (*and lawyers* in sentences with a complex filler NP and *those lawyers* in sentences with a simple filler NP. The maintenance site was defined as the relative clause region up until the RC verb (*who harshly* in sentences with SRCs and *who Andy harshly* in sentences with ORCs), and the retrieval site was defined as the RC verb *reprimanded* and the adverb of time *today*.

2.4.3 Results

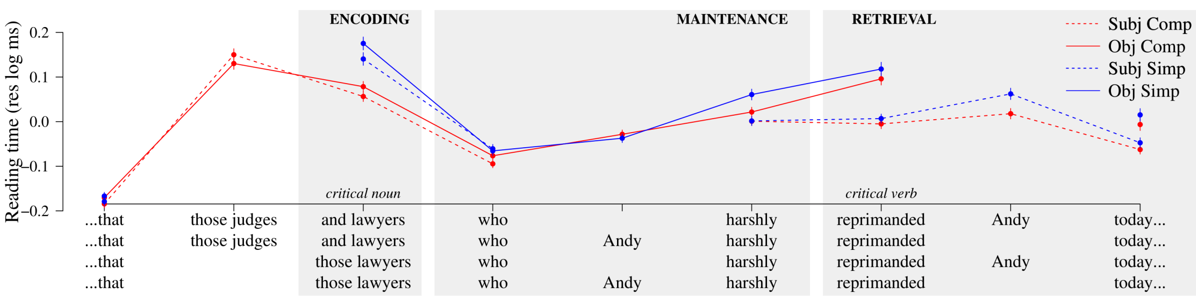

Figure 2.5: Experiment 2a: Residualized log reading times

In the **encoding site**, as evident in Figure 2.5 and Table 2.3, there was a main effect of filler NP complexity on the noun *those/and lawyers*. But the direction of the effect was the opposite to Experiment 1b: RTs decreased when the filler noun phrase was complex. This was also opposite to the findings in Hofmeister (2011) and Hofmeister and Vasishth (2014).

In the **maintenance site**, similar to Experiment 1a, we observed a main effect of RC type on the adverb *harshly*, where sentences with SRCs were read faster than those with ORCs. There was some weak evidence for a main effect of filler NP complexity on *who* and *harshly*, as well as an interaction between RC type and filler NP complexity on *harshly*. Consistent with Experiment 1a, participants tended to read sentences with a complex filler NP faster on these words than those with a simple filler NP. A post-hoc analysis on the word *harshly* also suggests that the facilitation effect of NP complexity was only reliable in ORCs but not in SRCs (Table 2.3), leading to the weak interaction on this word.

Region	**Effect**	**Mean**	**CrI lower**	**CrI upper**
those/and lawyers	Filler NP complexity	-0.091	-0.141	-0.041
who	Filler NP complexity	-0.022	-0.047	0.002
harshly	Filler NP complexity	-0.021	-0.045	0.004
	RC type	0.041	0.010	0.074
	Filler NP comp x RC type int.	-0.042	-0.093	0.008
	Filler NP complexity (SRC)	0.0006	-0.034	0.033
	Filler NP complexity (ORC)	-0.042	-0.079	-0.005
reprimanded	Filler NP complexity	-0.016	-0.041	0.009
	RC type	0.108	0.065	0.149
	Filler NP comp x RC type int.	-0.009	-0.061	0.042
today	Filler NP complexity	-0.018	-0.047	0.009
	RC type	0.058	0.027	0.090
	Filler NP comp x RC type int.	-0.005	-0.059	0.047

Table 2.3: Experiment 2a: Model summary for each region and fixed effect or interaction examined

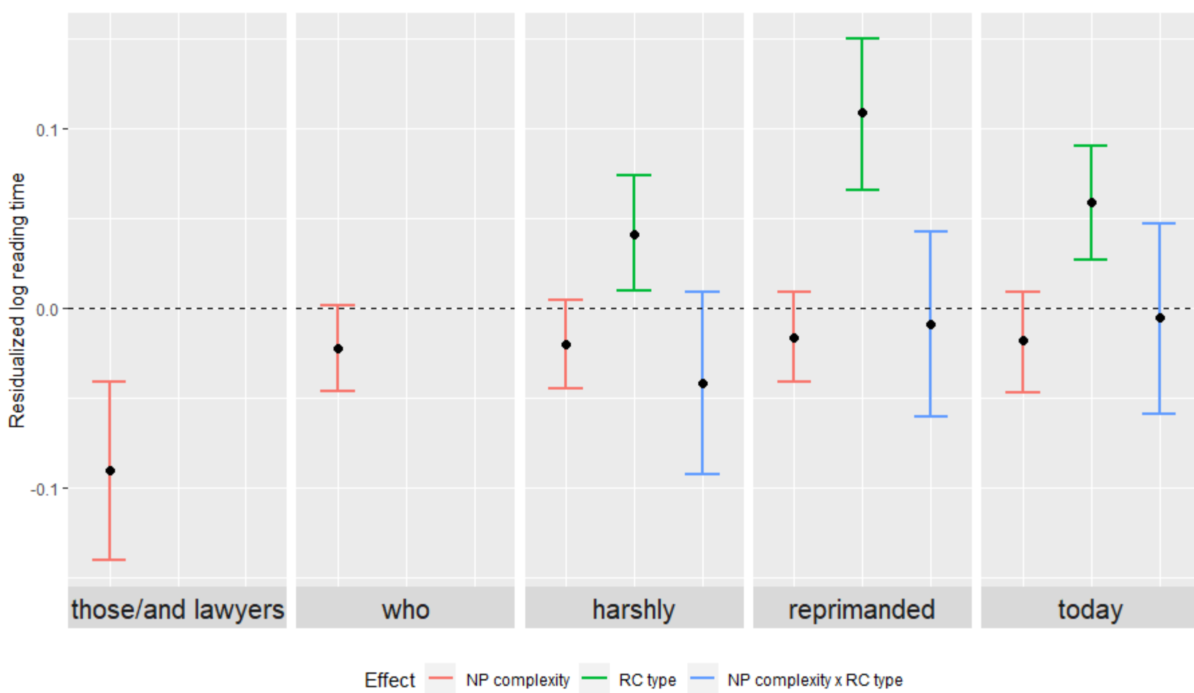

Figure 2.6: Experiment 2a: Credible intervals

In the **retrieval site**, similar to Experiment 1a, we found no speed up due to having a complex filler NP on the RC verb *reprimanded*. Neither did we find such an effect in any spillover words. As expected, we found the main effect of RC type on the RC verb *reprimanded* and the adverb *tonight* where the SRC advantage was evident.

2.4.4 Discussion

At the encoding site, the complex NP, instead of leading to a slowdown in RTs, resulted in faster RTs relative to the simple NP condition. This is different from what we observed in Experiment 1a/1b, and is also different from Hofmeister (2011) and Hofmeister and Vasishth (2014). This finding, however, should be treated with caution. There are a number of reasons why this observation may not directly speak to the issue of complexity effect in encoding. Firstly, at the encoding site that RTs were measured, the complex and simple conditions did not have exactly the same words. While the noun in this region was preceded by the demonstrative *those* in simple NPs, the noun is preceded by the coordinator *and* in complex NPs. Thus, the effect observed here can be due to the lexical differences between *those* and *and*. Secondly, the region is preceded by different words in the simple NP and the complex NP conditions. In sentences with simple NPs, the region is preceded by the complementizer *that*; and in sentences with complex NPs, the region is preceded by a content noun, e.g *judges*. Nevertheless, since previous word's RT was regressed out by residualization, this is unlikely to be the case. Lastly, in the complex but not the simple NP condition, the critical noun was preceded by another noun which, in most cases, was semantically related (*judges* and *lawyers*, *doctors* and *nurses*, etc). It is possible that the faster RTs we observed on the critical noun in the complex NP condition is due to semantic priming from the preceding noun.

At the retrieval site, we did not observe the complexity facilitation effect, unlike the findings in Experiment 1a and 1b, and also unlike the findings in previous work. We think this is a result of the fact that the complex NP condition was formed by conjoining two NPs together, whereas the complexity NP condition in Experiment 1a/1b and other studies was created by adding modifications on a noun. We will come back to this difference in the General Discussion section.

During maintenance, there is evidence, albeit somewhat weak evidence, that we replicated the complexity facilitation effect observed in Experiment 1a. Words in the maintenance site were read faster but only in ORCs, where there were two competing referents maintained in WM, *those judges and lawyers/ those lawyers* and *Andy*. This led to a weak Complexity x RC type interaction on the word *harshly* (see Figure 2.6 and Table 2.3). A notable difference between Experiment 1a and Experiment 2a on the word *harshly* is that the facilitatory effect was present for SRCs in Experiment 1a, but was totally absent in SRCs in Experiment 2a. This may again stem from the fact that the complex NP in Experiment 1a was a modified NP, whereas in Experiment 2a, it was a coordinated NP. We will come back to this observation in the General Discussion section.

2.5 Experiment 2b

2.5.1 Participants

100 participants aged 18-50 (mean age = 31.3), recruited through Prolific, participated in this experiment for payment. All participants self-identified as native, monolingual English speakers who were raised in monolingual households. In addition, participants had no language related disorders or literacy difficulties. Data from 16 out of 100 participants was removed because they scored below the threshold of 60% (1 standard deviation below the mean accuracy, rounded to the nearest multiples of 5%) for the comprehension questions.

2.5.2 Methods and Materials

Experiment 2b also had a 2 (NP complexity) x 2 (RC type) design. Parallel to Experiment 1a and 1b, Experiment 2b was similar to Experiment 2a except for the fact that an additional level of embedding *John thinks* was added to all conditions. The additional NP was always a proper name. An example set of stimuli for Experiment 2b is as follows:

(23) a. **Complex, SRC:** It seems / that / those judges / and lawyers / who / John / thinks / harshly / reprimanded / Andy / today / admitted / the error.

b. **Complex, ORC:** It seems / that / those judges / and lawyers / who / John / thinks / Andy / harshly / reprimanded / today / admitted / the error.

c. **Simple, SRC:** It seems / that / those lawyers / who / John / thinks / harshly / reprimanded / Andy / today / admitted / the error.

d. **Simple, ORC:** It seems / that / those lawyers / who / John / thinks / Andy / harshly / reprimanded / today / admitted / the error.

The additional level of embedding was also added to some fillers and to all comprehension questions. The experimental procedure was identical to the previous experiments.

The statistical analyses followed the same procedure as Experiment 2a. We removed 1.26% of reading time data that were 3 standard deviations of mean raw RT at each sentence and condition. For the regions of analyses, encoding and retrieval sites remained the same as Experiment 2a. The maintenance site now includes the new level of embedding *John thinks*.

2.5.3 Results

In the **encoding site**, from Figure 2.7 and Table 2.4, on the head noun *those/and lawyers*, reading times were faster for sentences with complex filler NPs, replicating the effect in Experiment 2a.

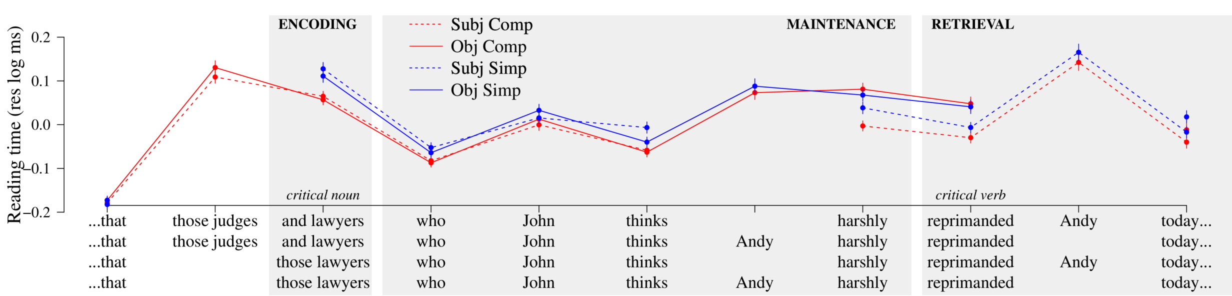

Figure 2.7: Experiment 2b: Residualized log reading time

In the **maintenance site**, there was some weak evidence for a main effect of filler NP complexity on *who*, with faster RTs on sentences that have a more complex head noun NP. This may be a spill-over effect from the previous region. On the verb *thinks*, by which point there were two competing referents in WM, we also observed weak evidence for a main effect of NP complexity, where participants read sentences with complex head nouns faster.

On the adverb *harshly*, there was a main effect of RC type, where sentences with SRCs

Region	**Effect**	**Mean**	**CrI lower**	**CrI upper**
those/and lawyers	Filler NP complexity	-0.062	-0.113	-0.011
who	Filler NP complexity	-0.026	-0.054	0.0004
John	Filler NP complexity	-0.019	-0.051	0.014
thinks	Filler NP complexity	-0.026	-0.053	0.0003
harshly	Filler NP complexity	-0.014	-0.042	0.015
	RC type	0.057	0.028	0.085
	Filler NP comp x RC type int.	0.052	-0.002	0.108
	Filler NP complexity (SRC)	-0.041	-0.078	-0.004
	Filler NP complexity (ORC)	0.012	-0.029	0.054
reprimanded	Filler NP complexity	-0.008	-0.037	0.021
	RC type	0.062	0.022	0.101
	Filler NP comp x RC type int.	0.022	-0.050	0.094
today	Filler NP complexity	-0.026	-0.059	0.008
	RC type	0.032	-0.005	0.067
	Filler NP comp x RC type int.	-0.007	-0.073	0.058

Table 2.4: Experiment 2b: Model summary for each region and fixed effect or interaction examined

were read faster. There was also weak evidence for an interaction between filler NP complexity and RC type. Visual inspection of Figure 2.7 showed that there seemed to be an effect of NP complexity but only in SRCs. A separate post-hoc analyses showed that the weak interaction was driven by the fact that the effect of NP complexity was only present for SRCs, but not ORCs. In SRCs, sentences with a complex head noun were read faster.

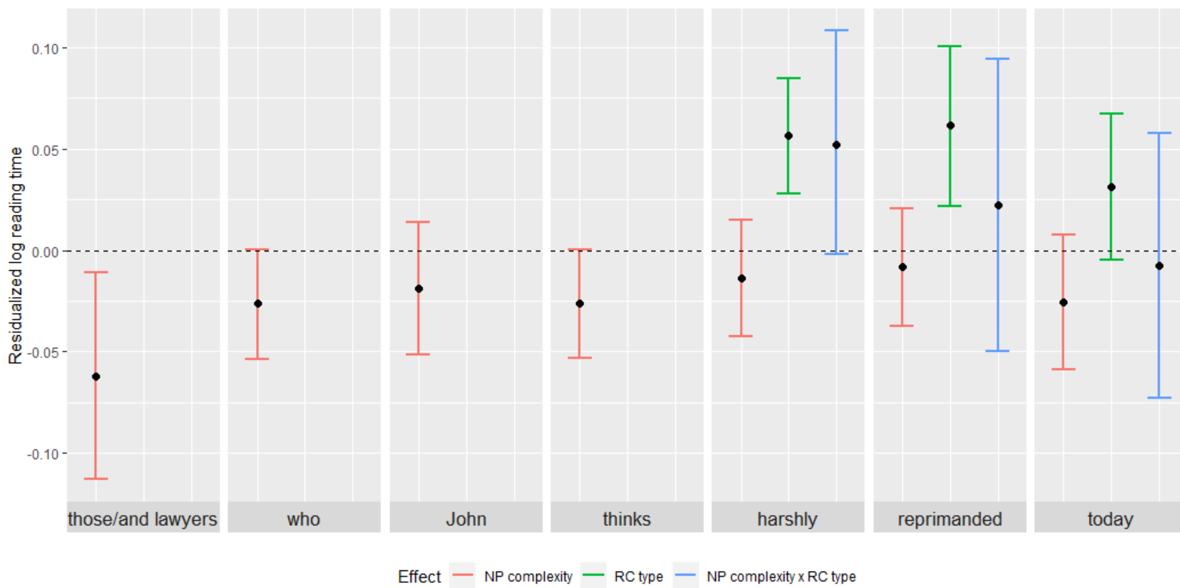

Figure 2.8: Experiment 2b: Credible intervals

In the **retrieval site**, there was no effect of NP complexity on the RC verb *reprimanded*, consistent with Experiments 1a and 2a. There was evidence for a RC type effect on the RC verb *reprimanded*, again with SRCs read faster than ORCs.

2.5.4 Discussion

The encoding site in Experiment 2b replicated the finding from Experiment 2a: there was a speed up in conditions with a complex filler NP relative to a simpler NP. Also consistent with Experiment 2a is that at the retrieval site and its spillover region, there was no facilitatory effect due to filler NP complexity. As discussed in section 4.4, this might be due to coordination not conferring the same set of processing benefits as modification.

At the maintenance site, there was weak evidence for a main effect of filler NP complexity at the words *who* and *thinks*. The speed up at *who* in sentences with complex filler NP might be the result of spillover from the encoding site. The speed up on the verb *thinks* is consistent with that in Experiment 1b. At this verb, there are two competing referents in WM: *those lawyers/ those judges and lawyers* and *John*. This result once again confirms our hypothesis that representational complexity helps maintain referents stored in WM more distinct and lowers maintenance cost.

A critical difference between the maintenance sites of Experiments 1b and 2b is the lack of the main effect of filler NP complexity on the adverb *harshly*. Instead, in this region, we observed weak evidence for the interaction between filler NP complexity and RC type, driven by facilitatory effect of filler NP complexity only in SRCs. Consistent with Experiments 1a and 2a, when there are two competing referents in WM (*those lawyers/ those judges and lawyers* and *John*) at the word *harshly* in SRCs, filler NP complexity facilitates processing by keeping the referents distinct from each other. However, when there are three competing referents in WM (*those lawyers/ those judges and lawyers*, *John* and *Andy*) in ORCs, the facilitatory effect of filler NP complexity disappears.

One possible explanation is that the processing cost to keep the representation of the target referent *those lawyers/ those judges and lawyers* distinct from competing referents was already incurred at the verb *thinks*. Since *Andy* and *John* are representationally similar as they are both human names, it is not necessary to modify representations to maintain the distinctiveness of *those lawyers/ those judges and lawyers* again. A post-hoc analysis of the ORCs in Experiment 1b showed that there was also no reliable effect of filler NP complexity on the same position. However, there was no interaction observed in Experiment 1b. In section 2.4.4, we discussed that factors contributing to the facilitatory effect of filler NP complexity other than maintaining distinctiveness appear to have a weaker effect when complexification is achieved via coordination than when it is achieved via modification. It is possible that in Experiment 1b's ORCs, the facitatory effect on *harshly* was7 mainly due to other factors such as increased salience. We will discuss more about this distinction between modification and coordination in the next section.

2.6 General Discussion

In four self-paced reading experiments, we investigated the benefit and cost of encoding, maintenance and retrieval of complex linguistic representations, with a particular focus on maintenance, which has been under-explored in the sentence processing literature.

For the encoding period, we found that only one of the two experiments using pre-modified NPs as complex NPs (Experiment 1b) replicated the slowdown due to the filler NP complexity found in Hofmeister (2011), despite both Experiments 1a and 1b having the same material up to and during the encoding period. The inconsistency between our two experiments suggests this effect might not be robust and might be subjected to inter-participant variation. Conversely, the two experiments using coordinated NPs as complex NPs (Experiments 2a and 2b) showed a facilitatory effect due to the filler NP complexity. Since the noun examined (such as *lawyers* is preceded by a related noun (such as *judges*) in a coordinated NP, it is likely that the noun in the encoding site is primed by the preceding noun, resulting in its increased accessibility and ease of encoding and as a result, faster RT.

On the other hand, the previously reported facilitatory effect during retrieval due to filler NP complexity (Hofmeister, 2011; Hofmeister and Vasishth, 2014; Karimi et al., 2020) was replicated in both Experiments 1a and 1b, on the RC verb and the spillover region. Contrary to experiments using pre-modified NPs, experiments using coordinated NPs, Experiments 2a and 2b, found no facilitatory effect due to filler NP complexity in neither the retrieval site nor its spillover regions.

Most critically, our results showed, for the first time, a facilitatory effect in the maintenance site when participants read sentences with a complex filler NP, no matter whether it is pre-modified or coordinated, suggesting a lower cost of maintaining featurally rich representations. We hypothesized that at least one possible source of this effect is due to the fact that rich semantic features on a representation helps to keep competing referent representations distinct from each other. Evidence supporting this hypothesis comes from SRCs and its comparison with ORCs. In Experiment 1a and 2a, when only the ORC but not the SRC construction contained competing referents in the maintenance site, NP complexity had a smaller (or even absent in Experiment 2a) facilitatory effect in SRCs than ORCs; but both constructions showed similar facilitatory effects or only SRCs showed facilitatory effect when an additional competing referent was added to the maintenance site in Experiments 1b and 2b, respectively.

We also observed that enhancing distinctness among competing referents is likely not the only reason that complex NP representations can facilitate the maintenance period. As we noted earlier, in Experiment 1a, even though there was a difference between SRCs and ORCs in terms of the facilitation effect brought about by the complex NP, there was nonetheless a reliable effect in SRCs. This effect goes beyond what our proposal about increasing distinctiveness can account for. As we noted in the Background, our explanation is not mutually exclusive with other explanations proposed by other researchers (Hofmeister, 2011; Karimi et al., 2020; Karimi and Ferreira, 2016). It is likely the additive effect of these multiple factors that generate the facilitatory effect of complex NP representations.

In addition to the main hypothesis we proposed for the maintenance period, we also consider two other possible interpretations of our results during the maintenance site. Firstly, the “maintenance" effect could have resulted from preemptive retrieval of the target NP in anticipation of the upcoming RC verb. In particular, on the preverbal adverb such as *dutifully* or the verb *thinks* in the additional clause *John thinks*, people may have anticipated that the RC verb is coming up, triggering a preemptive retrieval of the head noun. Since previous studies (Hofmeister, 2011; Hofmeister and Vasishth, 2014; Karimi et al., 2020) and our own Experiments 1a and 1b have demonstrated that retrieving a pre-modified NP from WM is easier, that effect might be extended to preemptive retrieval as well. However, this account alone does not explain the full pattern of our results. The effects observed on the preverbal adverb and the verb *thinks* did not match the effects observed on the RC verb itself in Experiments 1a, 2a and 2b. In the second set of experiments, 2a and 2b, which examined coordinated NPs, facilitatory effect was entirely absent on the verb and its spillover but was present on the preverbal adverb and the verb *thinks*.

The other possible interpretation is that our “maintenance" effect might actually be the result of memory encoding. Keeping referent representations distinct from each other can be beneficial for the encoding process. In fact, when similar items are encoded in WM, encoding interference could arise (Barker et al., 2001; Gordon et al., 2001; Hofmeister and Vasishth, 2014; Kush et al., 2015; Villata et al., 2018), due to the fact that similar items compete for the shared features and result in degraded representations of one or all items, a process known as feature overwriting (Nairne, 1990; Oberauer and Kliegl, 2006). In the current case, while the rich semantic features on the filler NP can give rise to higher encoding cost of this referent (at least for pre-modified NPs), subsequent encoding on later referents such as *Sophia* or *Jennifer* could be facilitated since the filler NP is more distinct from them. The challenge with this account is that it should have predicted the facilitation effect to arise on the nouns such as *Sophia* or *Jennifer*, at the moment when the encoding of the new referents took place. But in our results the facilitation effect thanks to filler NP complexity appeared after the proper names (e.g. on *dutifully* or *thinks*). The fact that the effect appeared after and not during the encoding of the new referents lends some support to our hypothesis that the cost of maintaining referent representations (that have already been encoded) could be reduced when competing referents have more distinct features.

2.6.1 Are All Types of Representational Complexity the Same?

In our experimental design, we examined two ways to make representations more complex, pre-modification and coordination, in order to investigate whether the effects observed can be generalized between different types of complex structures. The maintenance site is the only site that showed some similarities between the two sets of experiments. Both sets of experiments showed facilitatory effect due to complexity at this site (albeit to different degrees) and the distinction between SRCs and ORCs did largely align between them.

However, there were more differences than similarities between the two sets of experiments. In the encoding site, while we observed inconsistent inhibitory effect when the NP is pre-modified, we observed facilitatory effect instead when the NP was coordinated. In the retrieval site, we observed facilitatory effect only when the target NP was pre-modified (akin to that found in previous literature) but not when the target NP was coordinated. In the maintenance site, while the patterns were mostly similar, there were a few differences: 1) Unlike Experiment 1a, Experiment 2a using coordinated NPs showed no main effect of complexity on the preverbal adverb due to there not being a facilitatory effect due to complexity in SRCs; 2) Unlike Experiment 1b, Experiment 2b also showed no main effect of complexity on the preverbal adverb due to there not being a facilitatory effect due to complexity in ORCs.

While we have explained several hypotheses why a facilitatory effect was observed in the encoding site of Experiments 2a and 2b, it is possible that this discrepancy between pre-modified and coordinated NPs led to divergences downstream at the maintenance and retrieval sites. It is important to note that relative to Experiments 1a and 1b, Experiments 2a and 2b showed fewer effects at the maintenance and retrieval sites. The facilitatory effects on the preverbal adverb in SRCs in 2a and in ORCs in 2b, as well as on the RC verb in both 2a and 2b, were all absent compared to their equivalents in Experiments 1a and 1b. This can be the result of coordinated NPs conferring weaker benefits to processing in the maintenance and retrieval sites. However, it is unlikely that this is due to a weaker effect of increasing distinctiveness since as we noted, the overall patterning between ORCs and SRCs was preserved between the two sets of experiments. It is more likely due to the fact that encoding was facilitated, not inhibited, due to complexity in Experiments 2a and 2b. Hofmeister (2011) argued that one of the reasons that representational complexity is beneficial for the retrieval process is that the more effort needed to encode complex representations may increase the salience or activation level of the representation. Similarly, Karimi et al. (2020) hypothesized that the extended encoding time for longer NPs increased the attentional resources dedicated to them, making them easier to be retrieved. As a result, for our experiments, it seems that for coordinated NPs, the facilitatory effect during the encoding site must have conferred fewer attentional resources or less salience to the stored representation, leading to weaker or absent facilitatory effects during maintenance and retrieval.

This hypothesis suggests that the effects that were present in Experiments 1a and 1b and absent in Experiments 2a and 2b can be largely explained by increasing salience or attentional resources but not by increasing representation distinctiveness. This is in accordance with our hypothesis that multiple factors, beyond just increasing distinctiveness, explain the facilitatory effect of representational complexity. In fact, Karimi et al. (2020) found that increasing encoding time without adding semantic content also speeds up retrieval, supporting the conclusion that the facilitatory effect during retrieval can be largely explained by increasing salience and/or attentional resources.

There are, however, some caveats to the hypothesis that the time spent in the encoding site is what responsible for the discrepancies in the effects in the maintenance and retrieval sites between the two sets of experiments. Firstly, RT on the final word of the NP in the encoding site that we used for analysis can only be a proxy but not a comprehensive measure for encoding time or encoding effort. While the bulk of encoding into WM happens at this word, the measure should ideally take into account the time spent encoding the whole NP. However, if RT for the whole NP is considered, it is unclear how to normalize it in order to compare between the simple and complex conditions, as well as between the two sets of experiments. Secondly, it is important to reiterate that the slow down at the encoding site only occurred in Experiment 1a, not Experiment 1b, which is unusual given that the material was the same up to that point. Yet, they both showed facilitatory effect in the retrieval site. Further experimentation is needed to explore how different complex structures are encoded, maintained and retrieved in WM.

2.6.2 Connection to Encoding Interference

The facilitatory effect thanks to increased target NP complexity in the maintenance site that we observed is consistent with the hypothesis of encoding interference. When similar items are encoded in WM, they compete for the shared features and result in degraded representations of one or both items, a process known as *feature overwriting* (Nairne, 1990; Oberauer and Kliegl, 2006). In the sentence comprehension domain, researchers have hypothesized that feature overwriting degrades the representation of the target referent to be retrieved, leading to an increase in cost of retrieval and as a result, similarity-based interference at the retrieval site.

In our experiments, when the target NP is representationally complex, its feature richness keeps it distinct from competitor NPs. As a result, when a competitor NP enters WM, since the two NPs are less similar featurally, feature overwriting occurs to a lesser extent. While some shared features (such as both the target and competitor NPs being human nouns) are degraded in the representation of the target NP, the additional features (such as *emotional* and *crash*) are not degraded and thus, help maintain the high activation level of the target referent *emotional crash survivors*. Additionally, more encoding time for complex target referents means that more attentional resources are dedicated to them (Karimi et al., 2020). This also increases their activation level in WM. Since complex target NPs have higher activation level after feature overwriting than simple target NPs, less processing cost is needed to maintain them in WM. As a result, we observed faster reading time during the maintenance period for complex referent NPs.

2.7 Conclusion

In four experiments, we examined how the encoding, maintenance and retrieval of representationally complex NPs in WM differed from those of their simple counterparts, with a special focus on the understudied maintenance period. We found that storing more complex NPs reduces maintenance cost and that this effect can be partially attributed to the increased distinctiveness between the target NP and competing NP(s) in WM. Our results also replicated previous work that showed facilitatory effect of pre-modified representations on memory retrieval but this effect did not generalized to coordinated representations. We only found mixed evidence that representational complexity increases encoding effort for pre-modified NPs but decreases encoding effort for coordinated NPs.

CHAPTER 3

THE EFFECT OF REPRESENTATIONAL COMPLEXITY ON WORKING MEMORY PROCESSES - AN EEG PERSPECTIVE

3.1 Introduction

Chapter 2 demonstrated that varying the representational complexity of the noun phrase, either by modification with adjectives and nouns or by coordination, can affect the processing effort of the three WM processes. Most notably, the experiments showed, for the first time, facilitation of WM maintenance when the referent being maintained is representationally complex. While there are some previous works that provided behavioral evidence for effects during the maintenance period (Ness and Meltzer-Asscher, 2017,1; Ristic et al., 2021; Wagers and Phillips, 2014), much evidence for maintenance site effects was shown through neural measures such as EEG or fMRI (Fiebach et al., 2002,0; King and Kutas, 1995; Phillips et al., 2005). In this chapter, we corroborated our findings from the experiments in chapter 2 with results from Experiment 3, which examined, through an ERP lens, the effort associated with encoding, maintaining and retrieving a representationally complex NP.

3.1.1 What is Being Maintained?

Since the WM maintenance period is relatively understudied in the sentence comprehension literature compared to the retrieval period, there are still a lot of open questions when it comes to this period. Two major questions are still up for debate. Firstly, it is unclear whether referents are actively maintained in WM at all. This question was discussed in part in Section 2.1. Secondly, if referents are indeed actively maintained in WM, what exactly are being maintained, a subset of features of the referent or the referent in its entirety? Both of these questions will be explored in further details below.

The first question pertains whether referents are actively maintained in WM or are allowed to be decayed and then reactivated when they are needed during retrieval. This question was mainly explored through the examination of filler-gap dependencies, such as in the ORC in (24). In this example, upon encountering *who*, the comprehender realizes that they are reading a relative clause and thus have to carry the NP *the reporter*, called a filler, forward until the dependency is fulfilled. At the verb *attacked*, the filler is retrieved from WM and fills into the gap as the direct object/patient to complete the dependency. This type of dependency is thus named filler-gap dependency. The question here is whether comprehension system devotes resources to actively maintain the filler in WM or lets it decay prior to its retrieval.

(24) **The reporter** who the senator harshly attacked _ admitted the error.

The latter view is supported by a series of cross-modal lexical priming studies (Love and Swinney, 1996; Nicol and Swinney, 1989). In these experiments, participants listened to sentences that contained filler gap dependencies. However, while they were listening to certain words, they also had to perform the lexical decision task where they were prompted to indicate whether a probe string of letters they saw on the screen is a word in English or not. These probe words were manipulated to be either semantically related or unrelated to the filler. For example, the probe word for sentence (24) can be *news*, which is semantically related to *reporter* or *spoon*, which is semantically unrelated. The two critical positions where these probe words were presented were before or after the word *attacked*, the pre-gap and the gap positions, respectively. If the filler is actively maintained in WM, we would expect similar priming effects at the two positions. However, if the filler is left to be decayed, it is less activated at the pre-gap position while being reactivated in the gap position. Thus, we would expect the priming effect to be less pronounced or even absent at the pre-gap position. These priming studies found faster reaction times for the lexical decision task of semantically related words, but only at the gap position. These results are consistent with the view that the filler is not maintained at a high activation level until the gap position is encountered.

On the other hand, there is also behavioral evidence that suggests active maintenance of the filler. One piece of evidence comes from filled-gap effect such as that explored by the self-paced reading study of Stowe (1986). In sentence (25), the filler is *who* and the gap site is after the preposition *to*. However, there is a potential gap site after *bring* since readers have not yet read *us* at this verb. Stowe (1986) found a reading time slowdown at *us* as readers had to reanalyze their hypothesis. Frazier and d'Arcais (1989) argued that this slowdown reflected that the parser is eager to find a gap site for the filler and will rank the filler before lexical elements for gap analyses at predicates succeeding the filler, a process they called the *active filler strategy*. For this process to work, the filler has to be "not inert" in WM since it is considered at every predicate and thus, can be considered to be actively maintained.

(25) My brother wanted to know **who** Ruth will bring us home to at Christmas.

Some theorists (Ness and Meltzer-Asscher, 2017,1; Wagers and Phillips, 2014) tried to reconcile between these two views and seemingly conflicting sets of evidence. Wagers and Phillips (2014) proposed a middle ground where only some features of the filler are actively maintained in WM while others are left to decay and are reactivated at the retrieval site. The active maintenance of some but not all of the features of the filler is motivated by the limited capacity of WM and the need to reduce maintenance cost. This theory opened up the second question of which features are being actively maintained and which features are left to decay.

Wagers and Phillips (2014) investigated this using the filled-gap effect (Stowe, 1986). They tested the filled-gap effect in conditions where the filler is mismatched with the noun phrase selected by the verb in terms of syntactic category ((27) relative to (26)), semantic plausibility ((29) relative to (28)) and preposition subcategorization ((31) relative to (30)). In (27), *the chemicals* is not the direct object of *prepared* because the verb is looking for a NP, not a PP. In (29), *the cats* is a semantically anomalous direct object for *crammed*. In (31), *the courier* is not the direct object of *entrusted* because *entrusted* selects for the preposition *to*, not *from*. If the whole filler is actively maintained, we would expect similar extents for the filled-gap effect between these three different cases. While their experiments showed filled-gap effect happening immediately for all three cases outlined here, when the dependencies were lengthened by inserting a PP or Complementizer Phrase (CP) in between the filler and the potential gap, only (26)-(27) showed an immediate filled-gap effect while the filled-gap effects in (28)-(29) and (30)-(31) were delayed. For Wagers and Phillips (2014), this demonstrated that syntactic category information is actively maintained in a filler-gap dependency and thus exerts an immediate effect at the potential gap site no matter the dependency length. In contrast, other features such as the lexical semantics of the filler or the preposition identity are left to decay during maintenance and require more processing cost to retrieve when the dependency is longer. This theory reconciled the aforementioned discrepancy between the filled-gap effect results (Frazier and d'Arcais, 1989; Stowe, 1986) and the cross-model lexical priming results (Love and Swinney, 1996; Nicol and Swinney, 1989) since the experiment designs were probing different features of the filler: syntactic and semantic features, respectively. While there have been other attempts to determine what other features are maintained actively (Ness and Meltzer-Asscher, 2017,1), much is still unknown about the maintenance process.

(26) The chemicals which the technician carefully prepared the clean tubes for _ ...

(27) The chemicals for which the technician carefully prepared the clean tubes _ ...

(28) The acorns which the squirrels quickly crammed their small puffy cheeks with _ ...

(29) The cats which the squirrels quickly crammed their small puffy cheeks with _ ...

(30) The courier to whom the secretary warily entrusted the [...] correspondence _ ...

(31) The courier from whom the secretary warily entrusted the [...] correspondence _ ...

The SPR results from the four experiments in Chapter 2 showed that the representational complexity of the filler (whether the filler NP is pre-modified or not; whether the filler NP is coordinated or not) does have a facilitatory effect on reading time during the maintenance period. It is notable because this facilitatory effect occurred as late as the preverbal adverb, the word immediately prior to retrieval and it persisted even when the dependency length increases in Experiments 1b and 2b when a CP was added. The presence of the facilitatory effect late in the maintenance period despite the dependency length suggests that the representational complexity of the filler, like other syntactic features, is actively maintained in WM. To further investigate the active maintenance of the filler's representational complexity information in WM, we conducted an EEG experiment using stimuli similar to those of Experiment 1a, leveraging an ERP signal that has been argued to index maintenance of filler information: the Sustained Anterior Negativity (SAN).

3.1.2 Sustained Anterior Negativity

The SAN, also called the sustained left anterior negativity (SLAN), is a negative slow wave most pronounced in left frontal electrodes that occurs when a filler is being maintained in WM while awaiting for the gap. It has been detected in experiments comparing between contexts with a long filler-gap dependency and contexts with a short dependency or no dependency at all, such as ORCs relative to SRCs in English (King and Kutas, 1995), Mandarin (Yang et al., 2010), Japanese (Ueno and Garnsey, 2008) and Korean (Kwon et al., 2013); German object *wh*-phrases relative to subject *wh*-phrases (Fiebach et al., 2002), English object *wh*-phrases relative to *that*-phrases (Phillips et al., 2005), English *wh*- questions relative to *yes-no* questions (Cruz Heredia et al., 2022) and Japanese scrambled word order relative to unscrambled word order (Hagiwara et al., 2007; Yano and Koizumi, 2021). It also manifested in other contexts that are not filler-gap dependencies but demand the maintenance of linguistic materials in WM, such as coordinated NP relative to word lists (Lau and Liao, 2018) and temporal clauses with *before* relative to temporal clauses with *after* (Münte et al., 1998; Politzer-Ahles et al., 2017).

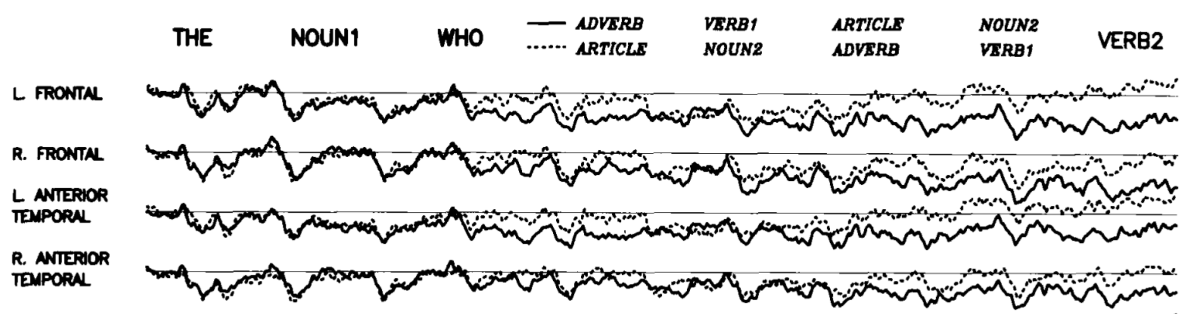

Figure 3.1: ORCs (dashed) showed SAN relative to SRCs (solid). SOURCE: King and Kutas (1995). Reproduced with permission from *MIT Press*.

The SAN was first described by King and Kutas (1995), who compared English ORCs (32) and SRCs (33). As seen in Fig 3.1, the frontal electrodes showed a negativity in ORCs relative to SRCs. This negativity persisted from the divergence point of the two RC types (the article in ORCs *the* and the adverb in SRCs *harshly*) past the end of the dependencies, the matrix verb *admitted*. Visual inspection of Fig 3.1 showed that the negativity was larger and started earlier in the left electrodes than the right electrodes.

(32) The reporter who the senator harshly attacked _ admitted the error.

(33) The reporter who _ harshly attacked the senator admitted the error.

Two functional interpretations of the SAN have been proposed, both centering around the role of WM (Cruz Heredia et al., 2022): *filler maintenance* and *syntactic prediction*. The *filler maintenance* interpretation aligns with the active maintenance hypothesis discussed in Section 3.1.1. The increased negativity is the manifestation of the increase in WM resources dedicated to maintaining the filler over a longer distance. For example, in the ORC (32), the filler *the reporter* has to be maintained over more words before encountering the gap after the RC verb *attacked* than in the SRC (32), where the gap occurs right after the complementizer *who*. Thus, it is more costly to maintain the filler in ORCs than SRCs, resulting in the SAN during the dependency. The alternative explanation, the *syntactic prediction* interpretation, aligns with the Dependency Locality Theory (Gibson, 1998), which was discussed in Section 2.1.1. For this interpretation, the increased negativity in the SAN results from having to maintain more predicted categories in WM. While both interpretations attribute the increased negativity to an increase in the amount of information being maintained, what exactly is being maintained - the filler or the predicted categories or both - remains a subject of debate. It is important to note that the presence of the SAN has been cited as supporting evidence for the active maintenance of the filler in WM (Ness and Meltzer-Asscher, 2019; Wagers and Phillips, 2014).

Following up on King and Kutas (1995), Fiebach et al. (2002), in juxtaposing object (34) and subject *wh*-phrases (35) in German, provided more data for better understanding of the SAN's functional interpretation. A drawback of King and Kutas (1995) was the significant lexical differences between the relative clauses. For example, the article *the* in the ORC (32) was compared to the adverb *harshly* in the SRC (33), making the interpretation of the SAN difficult since it is unclear how much of the negativity results from the lexical differences. Unlike English, German word order is relatively freer, allowing Fiebach et al. (2002) to compare between object and subject *wh*-phrases that have the same word order since German uses case morphology to mark grammatical subjects and objects. This manipulation significantly reduced the effect of lexical differences on the ERP signal. They also added a dependency length manipulation where *long dependency* conditions have three intervening PPs (*am Dienstag*, *nachmittag* and *nach dem Unfall*) as shown in examples (34) and (35) while *short dependency* conditions only have one intervening PP (*am Dienstag*). This manipulation was designed to examine the effect of dependency length on SAN amplitude and test the hypothesis that it gets progressively harder to maintain the filler as the dependency length increases.

(34) Thomas fragt sich, wen am Dienstag nachmittag nach dem Unfall
Thomas asks himself, who-ACC on Tuesday afternoon after the accident
der Doktor _ verständigt hat.
the-NOM doctor _ called has.
'Thomas asks himself who the doctor has called after the accident on Tuesday afternoon.'

(35) Thomas fragt sich, wer _ am Dienstag nachmittag nach dem Unfall
Thomas asks himself, who-NOM _ on Tuesday afternoon after the accident
den Doktor verständigt hat.
the-ACC doctor called has.
'Thomas asks himself who has called the doctor after the accident on Tuesday afternoon.'

As seen in Fig 3.2, Fiebach et al. (2002) did find a SAN in object vs subject *wh*-phrases but only in panel A when the dependency is long. A sustained signal is absent in panel B, where the dependency is shorter. The detection of the SAN by Fiebach et al. (2002) posed questions for the *filler maintenance* interpretation since the distance between the filler and the verb are the same in both (34) and (35) (Cruz Heredia et al., 2022). It lent support to the *syntactic prediction* interpretation since there were still more predicted categories in (34) than in (35).

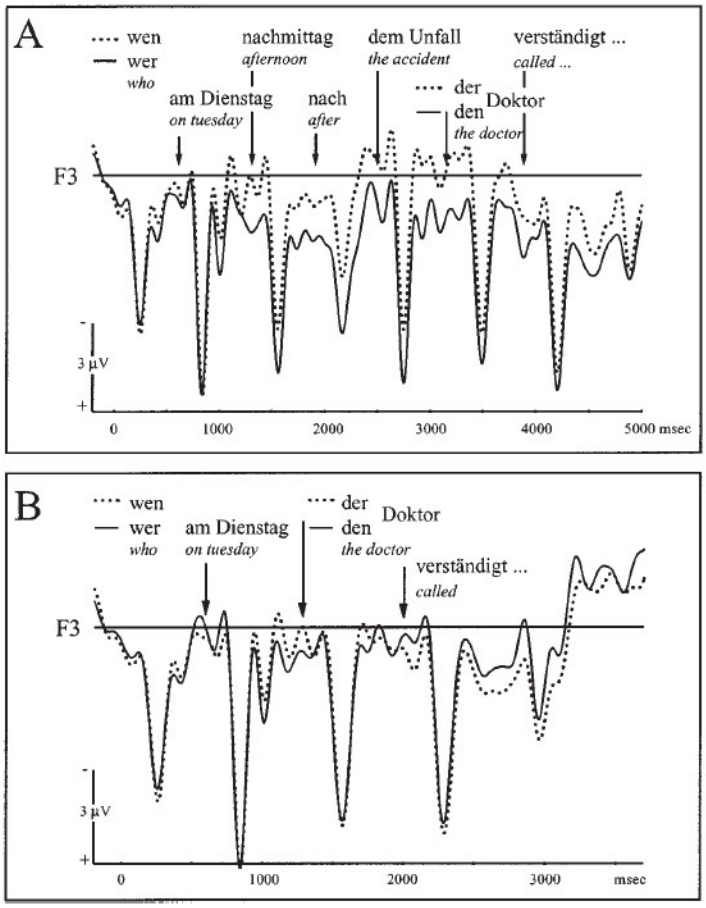

Figure 3.2: Object *wh*-phrases (dashed) showed SAN relative to subject *wh*-phrases (solid) but only in long dependency conditions (panel A). SOURCE: Fiebach et al. (2001). Reproduced with permission from *Elsevier*.

An additional finding from Fiebach et al. (2002) is that dependency length matters. Fiebach et al. (2002) hypothesized that the absence of the SAN during the short-dependency conditions is because of the much fewer memory resources needed for shorter dependency lengths. In addition, they also found that its amplitude and topographical distribution also grew with each successive word in the long dependency conditions, demonstrating progressively higher maintenance cost as the dependency gets longer. However, this increase in SAN amplitude along the dependency was not replicated when Phillips et al. (2005) compared *wh*-phrases (36) and *that*-phrases (37), as shown in Fig 3.3. They performed comparisons for each individual word's time window and found that later words in the dependency did not show an increase in SAN magnitude. Nevertheless, Phillips et al. (2005) still detected larger SAN in the long dependency conditions relative to short dependency conditions.

(36) The detective hoped that the lieutenant knew which accomplice the shrewd witness would recognize _ in the lineup.

(37) The detective hoped that the lieutenant knew that the shrewd witness would recognize the accomplice in the lineup.

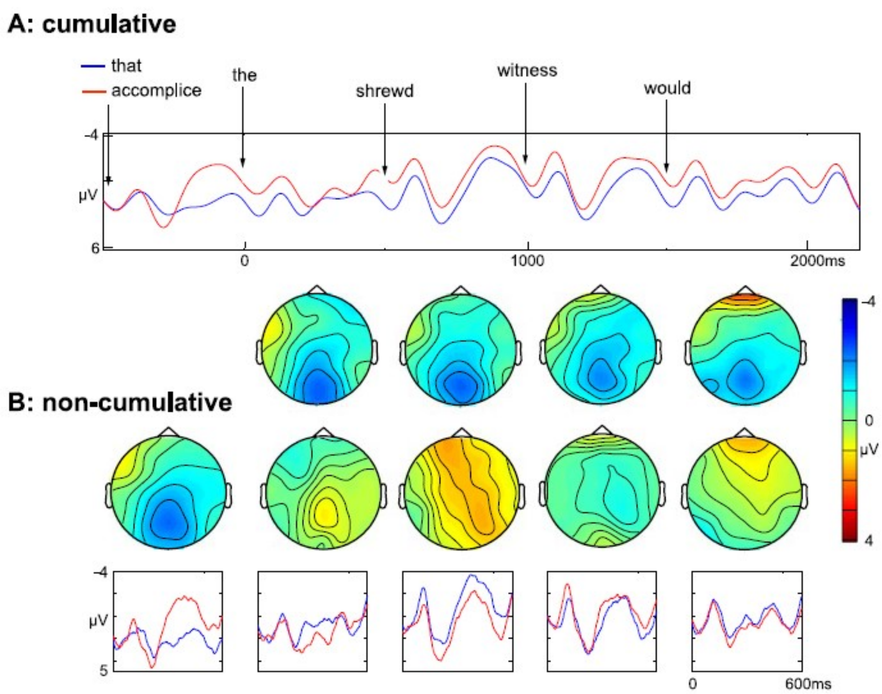

Figure 3.3: *wh*-phrases (red) showed SAN relative to subject *that*-phrases (blue) but SAN amplitude does not increase (panel B). SOURCE: Phillips et al. (2005). Reproduced with permission from *Springer Nature*.

While these foundational studies into the SAN supported WM-related interpretations of the SAN, whether it is *syntactic prediction* or *filler maintenance*, more recent studies have questioned these interpretations. Firstly, a study examining Japanese scrambled object-subject-verb (OSV) word order relative to the canonical non-scrambled subject-object-verb (SOV) word order (Yano and Koizumi, 2021) revealed that the presence of the SAN is modulated by pragmatics. They found that the SAN is present only when the OSV word order is not supported by context. Thus, it is unclear how to reconcile the WM-related interpretations of the SAN with its context sensitivity.

Furthermore, recent studies (Cruz Heredia et al., 2022; Sprouse et al., 2021) have failed to find the SAN in more memory-intensive contexts, which are predicted by the WM-related interpretations to generate the SAN. For instance, Cruz Heredia et al. (2022) found that matrix *wh*-dependencies (38) elicited the SAN relative to control (39) but embedded *wh*-dependencies ((40), relative to (41)) did not. Given that (38) and (40) have similar filler-gap dependencies, WM-related interpretations of the SAN cannot explain the discrepancy between them. They hypothesized that the SAN might not be syntax- or WM-related but instead reflect interpretive or pragmatic operations, which can explain the discrepancy between (38) and (40), as well as the context sensitivity of the SAN (Yano and Koizumi, 2021). Lau and Liao (2018) also proposed other possible interpretations of the SAN, that it might reflect subvocal rehearsal in the articulatory loop or implicit prosody processes in sentence comprehension.

(38) What did the commentary from the spokesman interrupt?

(39) Did the commentary from the spokesman interrupt the game?

(40) John wondered what the commentary from the spokesman interrupted.

(41) John wondered whether the commentary from the spokesman interrupted the game.

While more work is needed to fully understand the functional interpretation of the SAN, the majority of the existing results still point to its reflecting processes during WM maintenance and it is thus the closest index we have to the processing effort during maintenance process. It is also important to note that the idea of a sustained signal that indexes WM storage cost is not unique to sentence comprehension. Fuster and Alexander (1971) found persistent firing of some prefrontal cortex and thalamic neurons in monkeys when they had to maintain task-relevant materials across a delay period. Vogel and Machizawa (2004), using EEG to study visual WM, found a sustained ERP signal - the contralateral delay activity - whose amplitude is modulated by both the number of items in the memorandum and the WM capacity of the individual. Similarly, in the auditory domain, the sustained anterior negativity (which might or might not be related to the SAN in sentence comprehension) also correlates with the set size of pure tones participants are required to memorize through a delay period (Lefebvre et al., 2013; Nolden et al., 2013). As a result, in order to investigate how linguistic representations of differing complexity in maintained in WM, we still leverage understanding of the SAN while keeping the critiques against the WM-related interpretations of the SAN in mind.

3.1.3 The Current Study

The goal of Experiment 3 is to provide an EEG follow-up to the behavioral experiments in Chapter 2, especially focusing on the maintenance period thanks to the presence of the SAN. This allowed us to better understand how complex NPs are encoded and maintained in WM and retrieved from WM from both the behavioral and neural perspectives. In addition, we want to contribute to the discussion mentioned in Section 3.1.1 of what features of the filler are actively maintained in WM. The investigation can also give us additional data to contribute to better understanding the functional interpretation of the SAN.

As a follow up to the experiments in Chapter 2, the stimuli of Experiment 3 resembled closely those of Experiment 1a, with a few modifications, especially to make sure that lexical differences between the conditions in the same time window were minimized. If the representational complexity of the filler NP lowers WM maintenance cost, as we found in the behavioral experiments, we expect to detect the SAN between the simple NP conditions and the complex NP conditions during the maintenance period. For the retrieval process, we expect to detect a larger P600 effect for the simple conditions than for the complex conditions if having a complex filler NP facilitates retrieval. Previous works (Fiebach et al., 2002; Kaan et al., 2000; Phillips et al., 2005) have found larger P600 effects when there was increased cost of retrieving from WM and integrating into the current syntactic structure. Lastly, we want to explore possible ERP indices of encoding cost, which have not previously been described in the literature.

3.2 Experiment 3

3.2.1 Participants

52 participants (27, 20 and 5 identified as female, male and non-binary, respectively; 47 and 5 self-identified as right-handed and left-handed, respectively; age ranging between 18-28 years old, mean age = 20.04 years, standard deviation of age = 2.32 years) participated in this experiment for payment. All participants self-identified as native American or Canadian English speakers. In addition, participants had no language/attention/memory related disorders or literacy difficulties.

3.2.2 Stimuli

(42) **Complex, SRC:** It appears / that / the / merciless / prison / warden / who / today / harshly / reprimanded / Andy / admitted / the error.

(43) **Complex, ORC:** It appears / that / the / merciless / prison / warden / who / Andy / harshly / reprimanded / today / admitted / the error.

(44) **Simple, SRC:** It appears / that / the / warden / who / today / harshly / reprimanded / Andy / admitted / the error.

(45) **Simple, ORC:** It appears / that / the / warden / who / Andy / harshly / reprimanded / today / admitted / the error.

Sentences (42)-(45) are example sentences from the four conditions of the same sentence set. The forward slashes separate the words or groups of words that is presented individually. The experiment had a 2 x 2 design, where RC type (SRC/ORC) and filler NP complexity (complex/simple) were manipulated. The stimuli were designed to be similar in structure to those in Experiment 1a. Filler NP complexity was also manipulated by the inclusion or exclusion of two adjectives or nouns to modify the head noun. For complex conditions, the filler NP had two prenominal modifiers, such as *merciless* and *prison* in (42) and (45) above. The modifiers were chosen so that they were not improbable with the head noun (for instance, *kind dictator* would be improbable). All the modifiers and head nouns were unique between sentence sets. The RC type was manipulated by swapping the position between the RC noun *Andy* and the adverb of time *today* between SRC and ORC. This was done to ensure that the RC verb stayed at the same sentence position despite RC type (word 10 in sentences with complex filler NP and word 8 in sentences with simple filler NP). Unlike Experiment 1a, the sentence started with *It appears/seems that ...* so that the filler NP, a region of interest, was not at the beginning of the sentence.

Half of the sentence sets were followed by yes-no comprehension questions relevant to the sentence. The comprehension question was in the form of *Was it NP1 who {V-ed/was V-ed} by NP2?*, where NP1 could be the head noun of the filler NP without modifiers *the warden* or the RC noun *Andy* and NP2 was the other noun. Half of the comprehension questions were in passive voice so that participants could not rely on superficial positional cues to answer them. The other half of the questions were in active voice. The expected answers for half of the questions were *yes* and the expected answers for half of the questions were *no*.

200 sets of experimental stimuli formed four stimuli lists, created by a Latin square design. Thus, participants only read one sentence per sentence set. Each stimuli list was used for 13 participants.

3.2.3 Procedure

Participants performed two experiments back-to-back: this experiment and Experiment 5 detailed in Section 4.4. This experiment was always done first. The participant was assigned one of the four stimuli lists based on their participant number. Each trial consisted of: the fixation cross (1s), sentence presentation, comprehension question (50% of trials) and the *Ready?* screen. Sentences were presented using the RSVP paradigm, with each word or group of two words presented for 300 ms and an inter-stimulus interval of 200 ms. Participants were instructed not to move their eyes, face or body and not to read the sentence out loud during sentence presentation. In half of the trials, after the presentation of the last group of words of the sentence, a comprehension question appeared and participants had to press *f* if they thought the answer is *yes* and *j* if they thought the answer is *no*. For this experiment, there was no time limit for question answering. Participants were then encouraged to take a break and blink their eyes during the *Ready?* screen between trials and press the *Spacebar* to continue to the next trial when they are ready. The experimental trials (200 sentences) were randomly divided into five blocks (40 sentences per block). The experimenter provided feedback to the participant in between the blocks and encouraged them to take an extended break. The experimental trials were preceded by ten practice sentences, whose trial format was the same as that of the experimental trials.

3.2.4 Electrophysiological Recordings

EEGs were recorded from 32 electrodes from a Brain Vision actiCHamp Plus System located at Fp1/2, Fz, F3/4, F7/8, FC1/2, FC5/6, FT9/10, Cz, C3/4, T7/8, CP1/2, CP5/6, TP9/10, Pz, P3/4, P7/8, Oz, O1/2. A pair of electrodes (VEOG) was placed above and below the right eye and another pair (HEOG) was placed at the outer canthi of both eyes to monitor vertical and horizontal eye movements respectively. EEGs were referenced online to the average of all EEG electrodes. Impedances of all EEG electrodes were maintained below 5 kΩ for the entire duration of the experiment. Continuous data were digitized at a sampling rate of 1000 Hz.

3.2.5 Data Preprocessing and Analysis

Preprocessing of the raw data was done using EEGLAB (Delorme and Makeig, 2004) and ERPLAB (Lopez-Calderon and Luck, 2014), which are extensions of MATLAB. EEG data were band-pass filtered between 0.01 and 40 Hz, then re-referenced to the average of the two mastoid electrodes TP9 and TP10.

Since we are interested in the EEG signal that corresponds to the encoding and maintenance period, as well as the retrieval period, we epoched the EEG data into three epochs. Epoch 1, which reflected the period where the filler NP was encoded into WM and maintained there, was a 2500-ms epoch starting from the onset of the head noun of the filler NP *warden* and lasting until the offset of the blank screen after the RC verb *reprimanded.* Epoch 2 was a 2500-ms epoch that overlapped with most of Epoch 1 but started from the onset of the complementizer *who* and lasted until the offset of the blank screen after *Andy/today*. Analysis on Epoch 2 was done because the words immediately pre-baseline in Epoch 1 were different between the complex filler NP and simple filler NP conditions, possibly giving rise to ERP signals that were artifactual. In Epoch 2, however, the word immediately prior to the baseline was the noun *warden* in all conditions. To avoid repetition, I only reported the filler NP complexity comparison for Epoch 2. Epoch 3, which reflected the period where the filler NP was retrieved from WM, was a 2000-ms epoch starting from the onset of the RC verb *reprimanded* and lasting until the end of the sentence. The data in all epochs were baselined by the 100 ms immediately before the epoch. For artifact rejection, we carried out peak-to-peak thresholding of the VEOG and HEOG electrodes (200 ms time window with 100 ms step) to exclude trials whose changes were more than 70 mV. We also removed trials with extreme voltages in EEG electrodes that were not within the range of [-75 mV, 75 mV]. Additionally, for each epoch individually, we excluded epochs from any participants whose rejection rate was more than 50%. After this procedure, data from 44 out of 52 participants were retained for Epoch 1 and 2 and data from 45 participants were retained for Epoch 3. We removed 35.4%, 36.1% and 33.2% of trials from Epoch 1, 2 and 3, respectively through this artifact rejection procedure. There were 37.8, 37.4 and 38.3 trials left on average per condition for Epoch 1, 2 and 3, respectively.

To detect reliable differences for both manipulations (RC type and filler NP complexity) between the ERPs, we used a repeated measures, two-tailed cluster-based permutation test based on the cluster mass statistic (Bullmore et al., 1999) using a family-wise alpha level of 0.01. This was done using the Mass Univariate ERP Toolbox (Groppe et al., 2011), an extension of MATLAB. The data was first downsampled from 1000 Hz to 250 Hz using boxcar filter and then re-baselined using the 100 ms immediately before the epoch. All time points after downsampling (625 time points for Epochs 1 and 2 and 500 time points for Epoch 3) and all 32 EEG electrodes were used in the test (i.e, 20000 comparisons for Epoch 1 and 2 and 16000 comparisons for Epoch 3). Repeated measures t-tests were performed for each comparison for the original data, as well as 2500 random within-participant permutations of the original data. For each permutation, t-scores that corresponded to uncorrected p-values of 0.01 or less were combined with neighboring t-scores that also fulfilled the same condition to form clusters. Spatial neighbors were defined as electrodes whose distance was no more than 5.77 cm. Temporal neighbors were adjacent time points. We then calculated the mass of all the clusters - the sum of t-scores in that cluster. The most extreme cluster mass (whose absolute value was the largest) for each of the 2501 permutations was recorded, which together formed the null distribution. Using this distribution, p-values were derived for each cluster in the original data, allowing us to derive statistically significant clusters. Two tests were carried out for each epoch using this procedure, one for the difference between sentences with ORCs and SRCs and one for the difference between sentences with simple and complex filler NP.

Additionally, we used the Factorial Mass Univariate ERP Toolbox (FMUT) (Fields, 2017) to examine the interaction between the complexity of the filler NP and RC type. Fields and Kuperberg (2020) argued that the MUT analysis is only applicable for simple designs and FMUT was designed to make use of mass univariate approaches for factorial ANOVA to investigate complex factorial designs. Our preprocessing procedure prior to the statistical test for the factorial cluster-based permutation test using FMUT was the same as that for the cluster-based permutation test using MUT. The test was also carried out using a family-wise alpha level of 0.01.

3.2.6 Results

For the 2500-ms Epoch 1, the epoch that corresponds to the encoding and maintenance of the filler NP into WM, visual inspection of Fig 3.5 demonstrated that there was a sustained negativity between the simple conditions and the complex conditions that lasts for the whole epoch. This sustained negativity was global and has similar magnitude accross the different electrodes, except for the right frontal electrode F4, which showed much a much smaller negativity. Moreover, the magnitude of the negativity did not appear to change over time through the epoch. The topogram Fig 3.6 confirmed the presence of this negativity as it included a largely global sustained negativity in both SRCs and ORCs. The outcome of the cluster-based permutation test supported the visual inspection as there was a main effect of filler NP complexity for the whole time window across one significant cluster where conditions with complex filler NP showed more negativity. The most consistent effect came from frontal, fronto-central, central, centro-parietal and parietal electrodes that surround the midline.

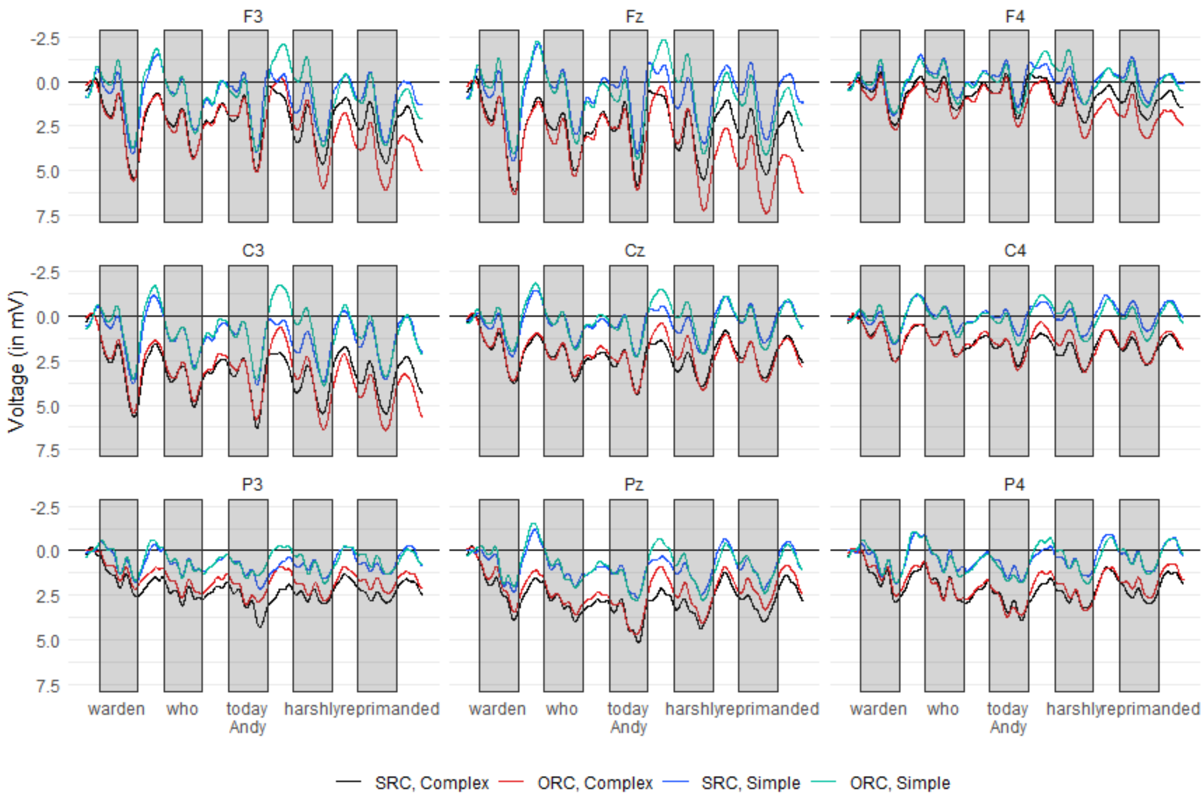

Figure 3.4: Experiment 3: Grand averaged ERPs for Epoch 1 of all conditions

Regarding the RC type comparison in Epoch 1, visual inspection of Fig 3.7 revealed a transient difference starting 300 ms (200 ms for electrode P3) after the onset of *today/Andy*, the site where the words diverged between the ORC and SRC conditions. The ORC conditions were more negative than the SRC conditions at the midline and left electrodes and this negativity last until 100 ms after the onset of the next word, *harshly*. Additionally, there was a sustained frontal positivity between ORCs and SRCs that started 300 ms after the onset of the adverb *harshly* and grew larger in magnitude, as evident in Fig 3.8, especially in the complex conditions.

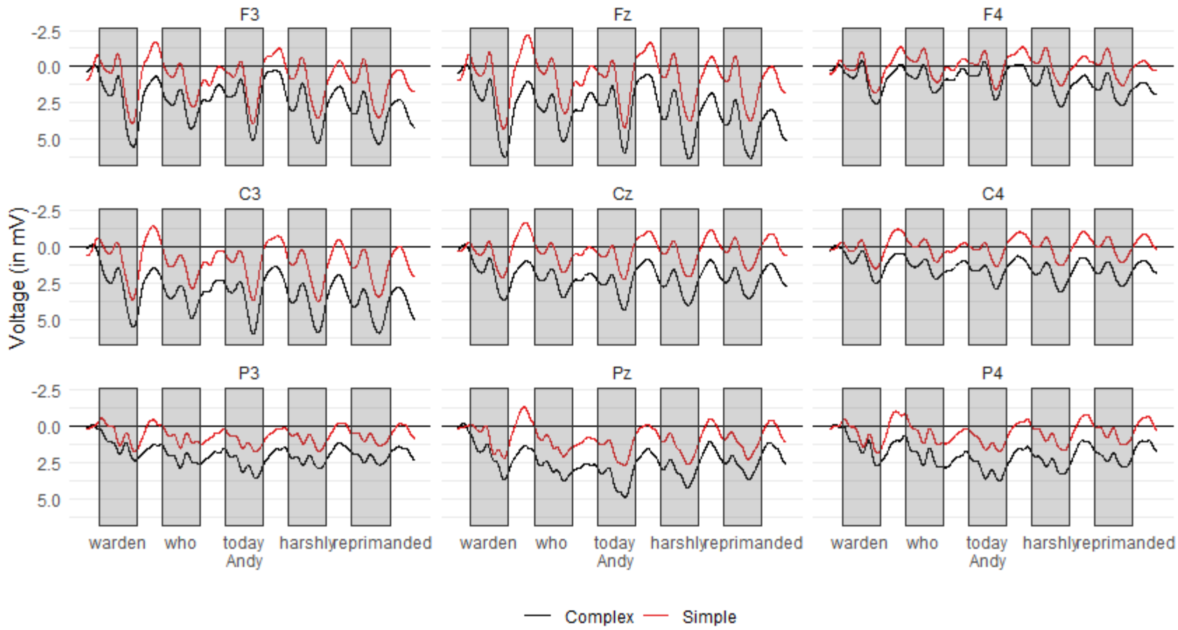

Figure 3.5: Experiment 3: Grand averaged ERPs by filler NP complexity for Epoch 1

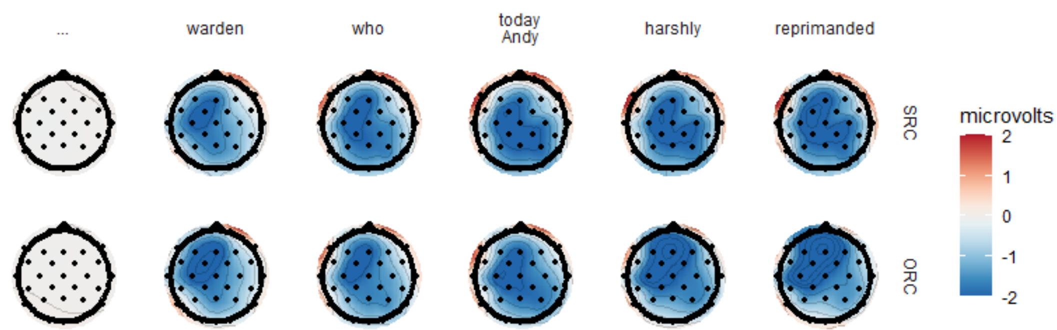

Figure 3.6: Experiment 3: Topograms for Simple-Complex at both SRC and ORC conditions for Epoch 1

The outcome of the cluster-based permutation test had three significant clusters: one positive cluster and one negative cluster in the span of 1080-1616 ms and one positive cluster in the span of 1660-2496 ms. In the span of 1080-1616 ms, most left and midline electrodes were more negative in ORCs while most right electrodes were more positive. The positive cluster from 1660-2496 ms, which showed significance in the prefrontal, frontal, fronto-central and temporal electrodes on both hemispheres, reflected the sustained positivity observed through visual inspection.

Post-hoc permutation test involving only complex filler NP conditions revealed a significant positive cluster from 1672 ms to the end of the epoch. This confirmed the sustained positivity in ORC conditions relative to SRC conditions when the NP was complex. On the other hand, no significant cluster was detected for the same period of time for the simple filler NP conditions. This discrepancy supported our earlier observation through visual inspection that the sustained frontal positivity was more prominent in the complex conditions. No significant interaction between filler NP complexity and RC type was detected.

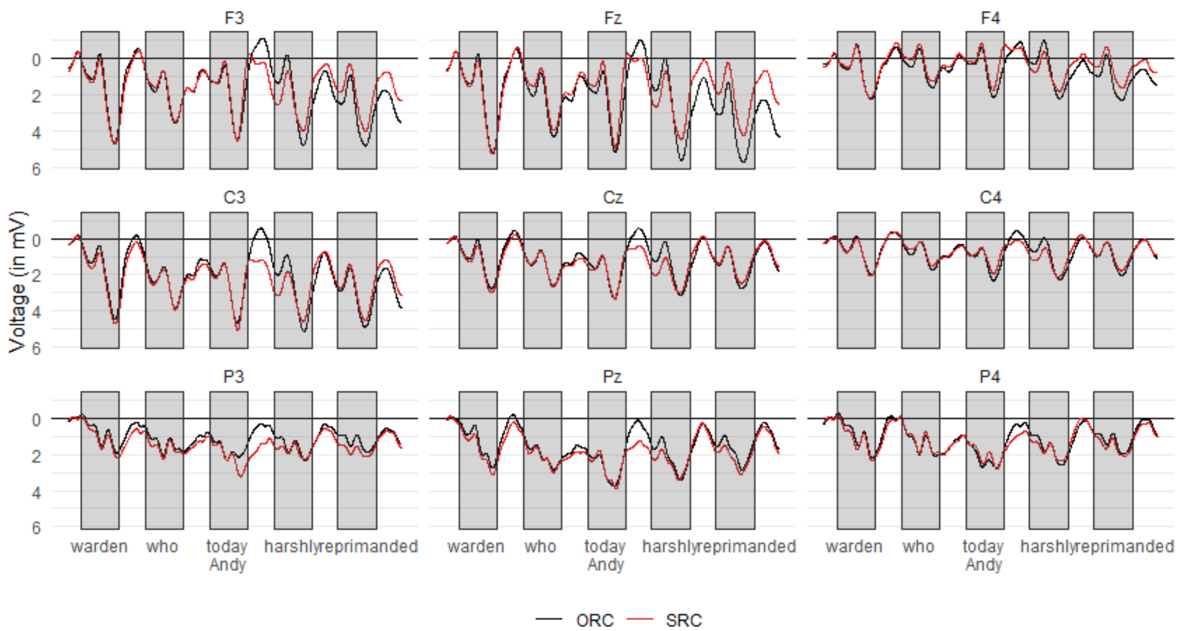

Figure 3.7: Experiment 3: Grand averaged ERPs by RC type for Epoch 1

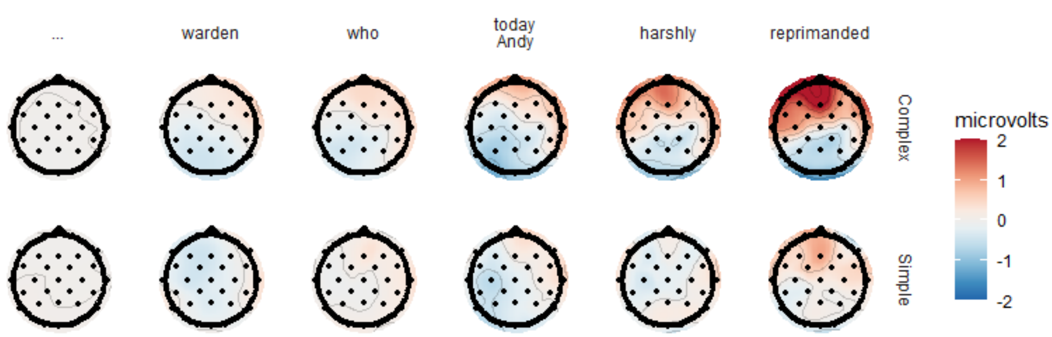

Figure 3.8: Experiment 3: Topograms for ORC-SRC at Complex and Simple conditions for Epoch 1

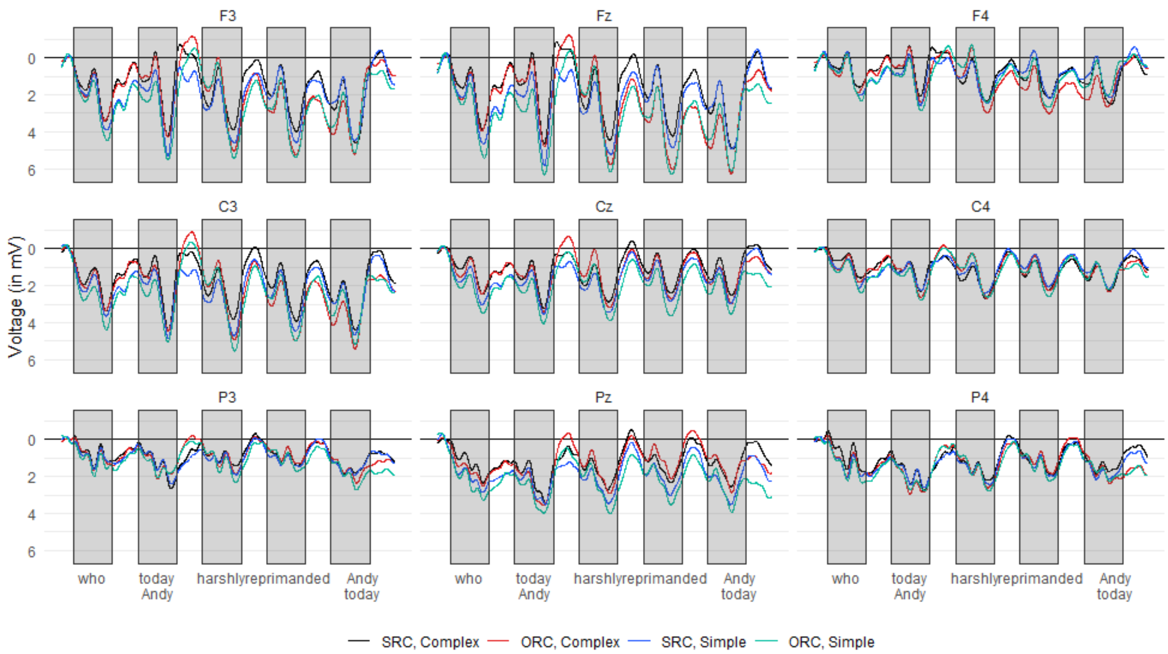

Figure 3.9: Experiment 3: Grand averaged ERPs for Epoch 2 of all conditions

For the 2500-ms Epoch 2, which corresponds to the maintenance and retrieval of the filler NP, we only reported the result of the filler NP complexity comparison since Epoch 2 overlapped with most of Epoch 1. Epoch 2 was analyzed because it was unclear whether the sustained global negativity observed in the filler NP complexity comparison in Epoch 1 was due to a pre-baselining artifact. Unlike Epoch 1, the words immediately prior to Epoch 2 were the same across conditions.

Visual inspection of Figs 3.10 and 3.11 revealed that the simple NP conditions elicited more positivity than the complex NP conditions through the first three words of the epoch, *who*, *today/Andy* and *harshly*. This positivity were largely left frontal and decreased in magnitude gradually. The permutation test showed 2 positive clusters: one spanning 116-504 ms (on the word *who*) and one spanning 720-932 ms (on the word *today/Andy*). Both clusters concentrated around left frontal electrodes, confirming the sustained left anterior positivity elicited in simple filler NP conditions relative to complex filler NP conditions during visual inspection.

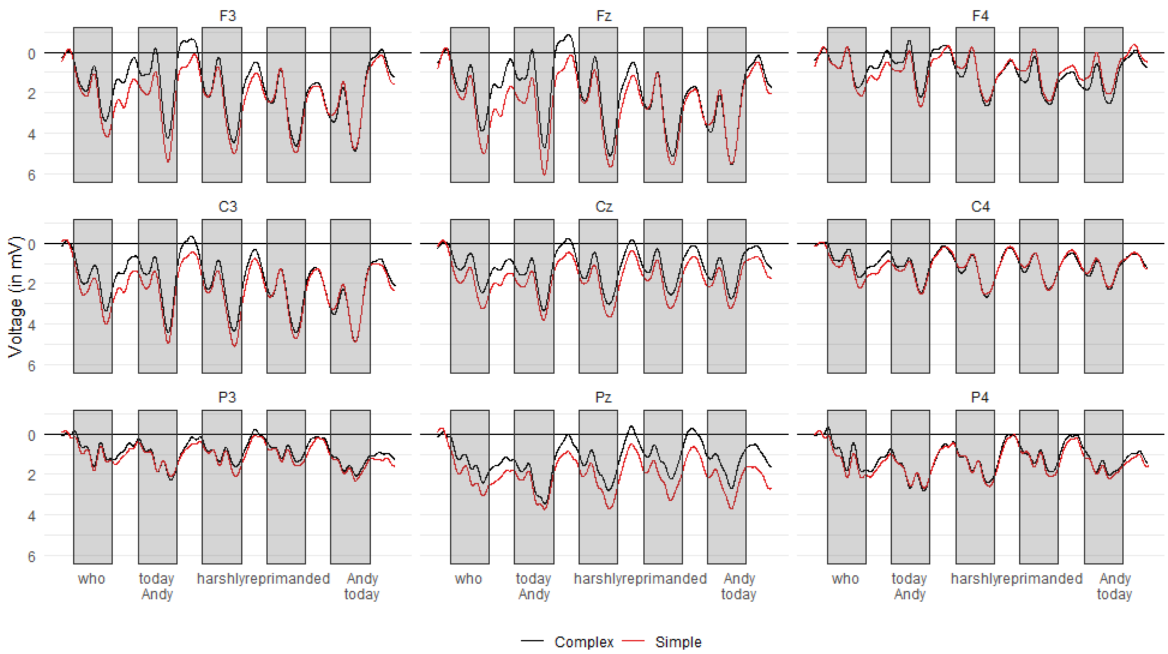

Figure 3.10: Experiment 3: Grand averaged ERPs by RC type for Epoch 2

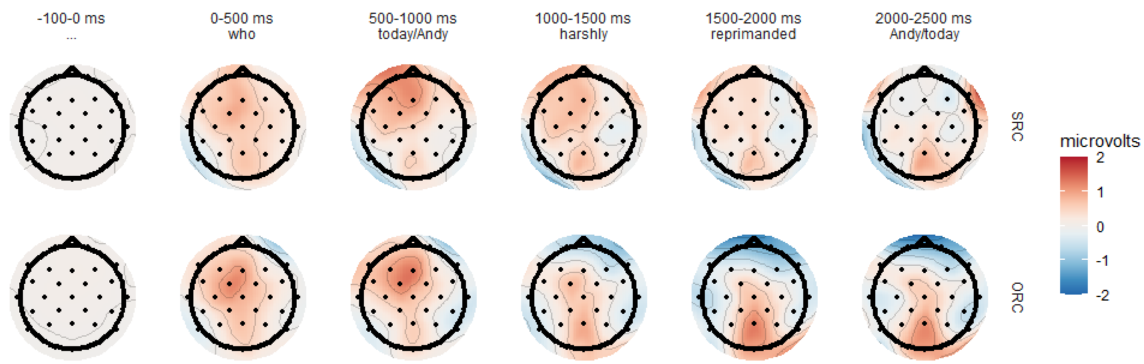

Figure 3.11: Experiment 3: Topograms for ORC-SRC at Complex and Simple conditions for Epoch 2

For the 2000-ms Epoch 3, which corresponds to the retrieval of the filler NP from WM, visual inspection of Fig 3.13 showed only small transient frontal positivities between the simple and complex conditions 300ms after the final two words, *admitted* and *the error*, which can also be seen in Fig 3.14. However, no significant cluster was found in the cluster-based permutation test.

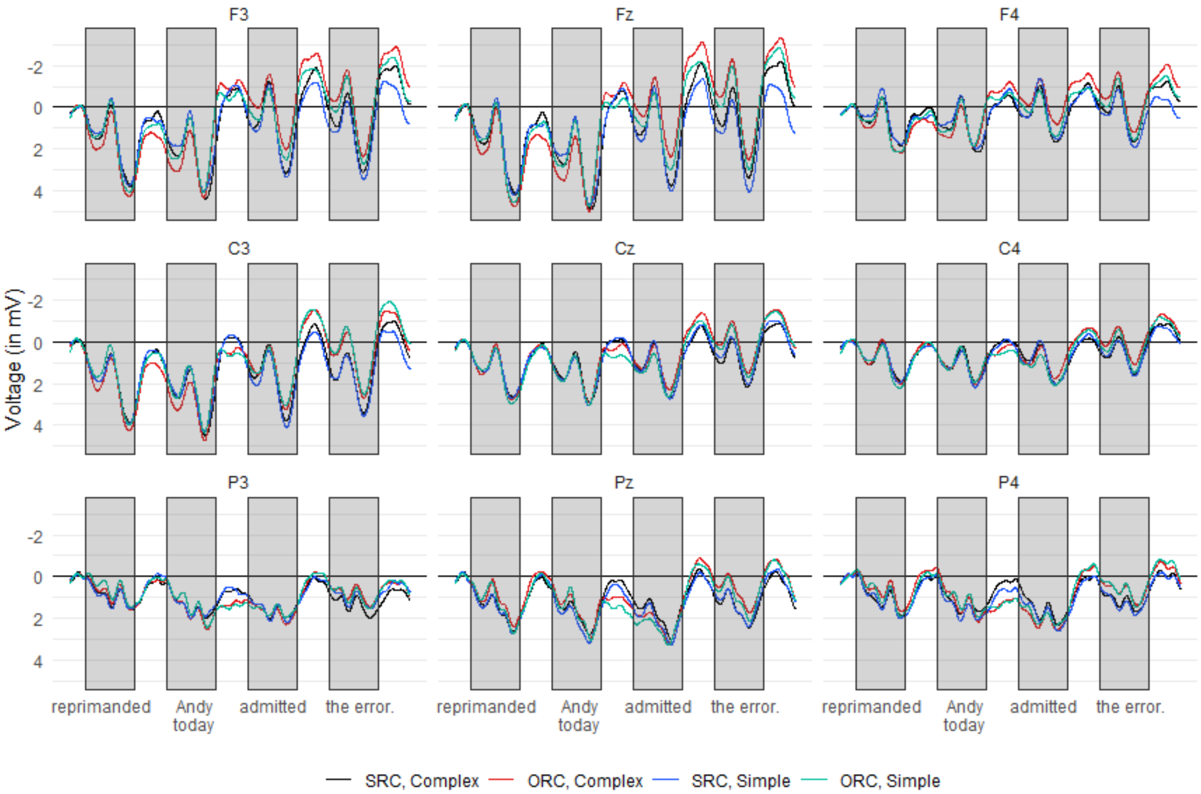

Figure 3.12: Experiment 3: Grand averaged ERPs for Epoch 3 of all conditions

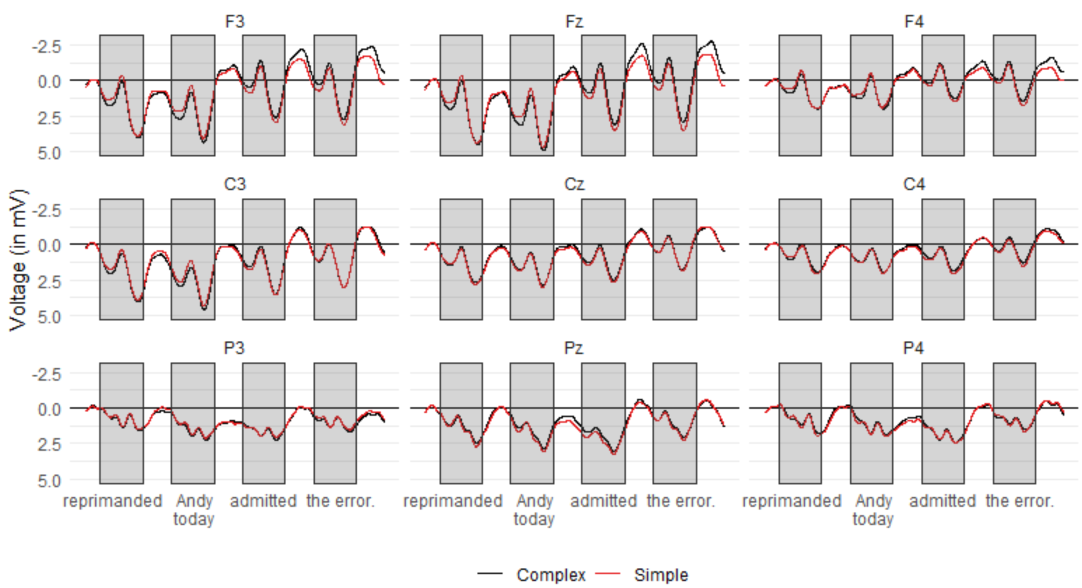

Figure 3.13: Experiment 3: Grand averaged ERPs by filler NP complexity for Epoch 3

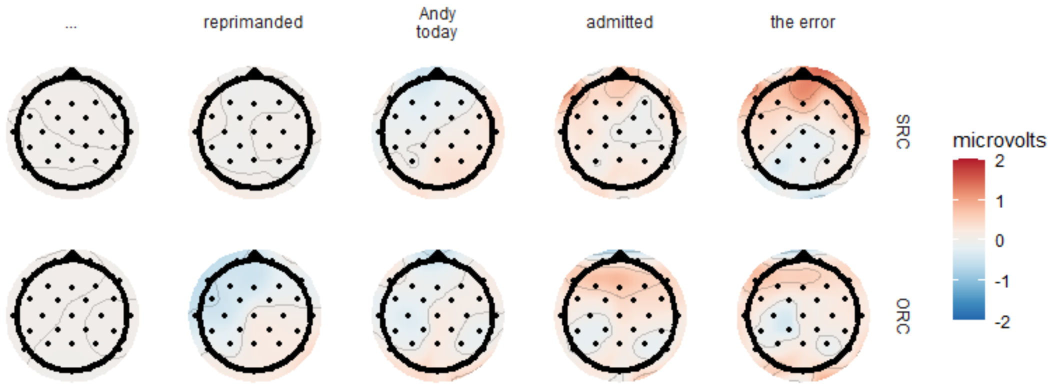

Figure 3.14: Experiment 3: Topograms for Simple - Complex at SRC and ORC conditions for Epoch 3

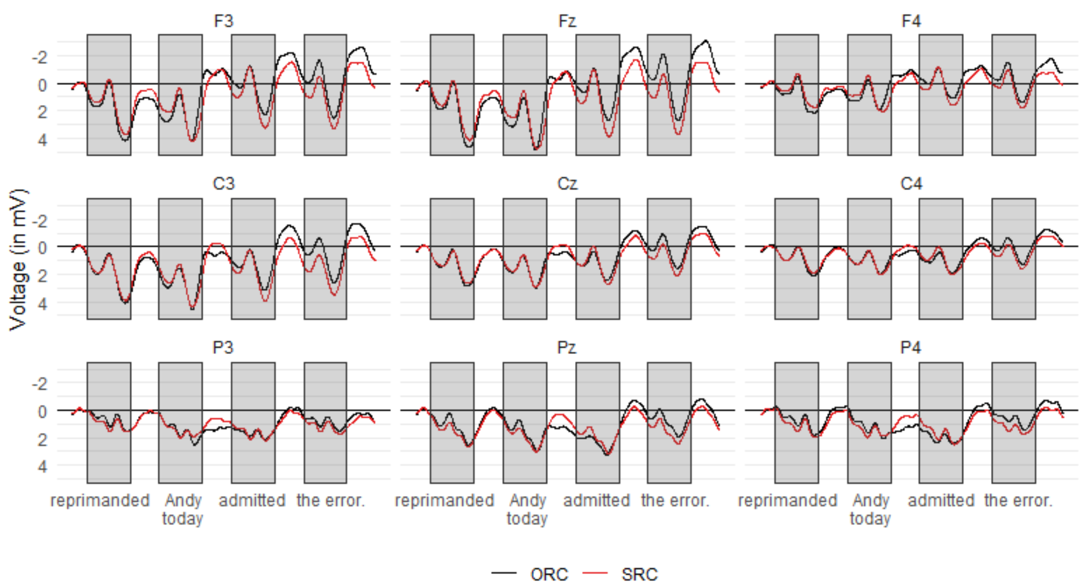

Figure 3.15: Experiment 3: Grand averaged ERPs by RC type for Epoch 3

For the RC type comparison in Epoch 3, visual inspection of Fig 3.15 revealed a transient positivity in ORCs compared to SRCs in the central and parietal electrodes 300 ms after *Andy/today*, the site where the two RC types diverged. This positivity only last about 200 ms. There was also a sustained negativity that was largely left frontal (F3, Fz and C3 in Fig 3.15) spanning from 300 ms after the filler verb *admitted* to the end of the epoch. As can be seen in Fig 3.16, this negativity was similar in magnitude and distribution between the complex and simple conditions.

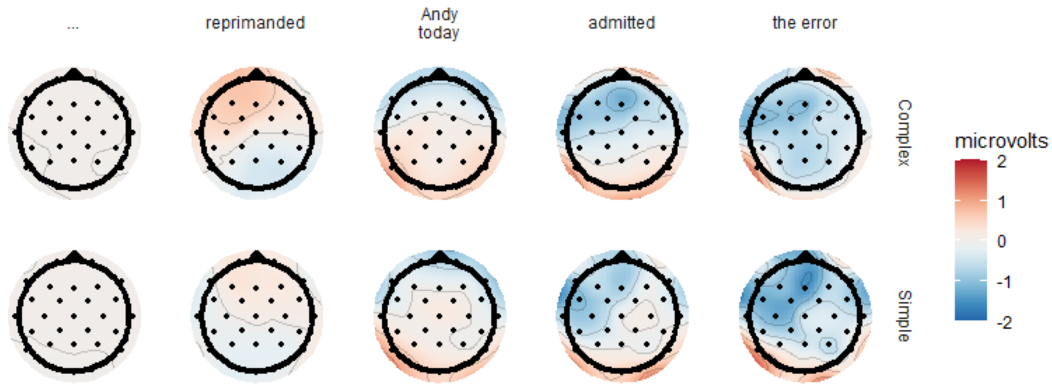

Figure 3.16: Experiment 3: Topograms for ORC-SRC at both Complex and Simple conditions for Epoch 3

The permutation test revealed three significant clusters: one positive cluster and one negative cluster in the span of 660-1144 ms and one negative cluster in the span of 1140-1996 ms. In the span of 660-1144 ms, ORC conditions elicited more negativity in frontal and fronto-central electrodes and more positivity in posterior electrodes than SRC conditions. In the span of 1140-1996 ms, ORC conditions elicited more negativity than SRC conditions, particularly on left fronto-central electrodes. This cluster reflected the left sustained anterior negativity observed during visual inspection. Akin to Epoch 1, no significant interaction between filler NP complexity and RC type was detected in Epoch 3.

3.3 General Discussion

In this experiment, our goal was to investigate through an ERP lens the effect of the filler NP's representational complexity on its encoding, retrieval and especially maintenance. This examination can also inform us whether representational complexity is among the features that are actively maintained in WM in a filler-gap dependency. Additionally, the presence or absence of the SAN in this experiment, which also included an RC type manipulation, can contribute to the discussion about the functional interpretation of the SAN.

Epoch	Effect	Timing	Electrode
1	Filler NP complexity (1 cluster)	0-2496 ms	FP1, Fz, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, Pz, P3, P7, O1, Oz, O2, P4, P8, TP10, CP6, CP2, Cz, C4, T8, FT10, FC6, FC2, F4, F8, FP2
	RC type (2 clusters)	1080-1616 ms	FP1, Fz, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, Pz, P3, P7, O1, Oz, O2, P4, P8, TP10, CP6, CP2, Cz, C4, T8, FT10, FC6, FC2, F4, F8, FP2
		1660-2496 ms	FP1, Fz, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, P3, P7, O1, Oz, O2, P4, P8, CP6, Cz, C4, T8, FT10, FC6, FC2, F4, F8, FP2
2	Filler NP complexity (2 clusters)	116-504 ms	FP1, Fz, F3, F7, FC5, FC1, C3, CP1, Pz, Oz, O2, P4, P8, CP2, Cz, C4, FC6, FC2, F4, FP2
		720-932 ms	FP1, Fz, F3, F7, FT9, FC5, FC1, C3, Cz, FC2
3	RC type (2 clusters)	660-1144 ms	FP1, Fz, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, Pz, P3, P7, O1, Oz, O2, P4, P8, CP6, CP2, Cz, C4, T8, FT10, FC6, FC2, F4, F8, FP2
		1140-1996 ms	FP1, Fz, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, Pz, P3, P7, O1, Oz, O2, P4, P8, TP10, CP6, CP2, Cz, C4, T8, FT10, FC6, FC2, F4, F8, FP2

Table 3.1: Experiment 3: Timing and electrodes of significant clusters from FMUT analysis

3.3.1 On Representational Complexity

For the filler NP's representational complexity contrast, we manipulated the number of pre-modifiers of the filler NP. Epoch 1, where the filler NP is encoded into WM, maintained there and starts to be retrieved, showed that at all electrodes on and around the midline, the simple NP conditions elicited more negativity than the complex NP conditions throughout the whole epoch. This sustained negativity aligns with our hypothesis that having a representationally complex NP facilitates WM maintenance of the NP filler in the dependency.

Since representationally complex NPs are more distinctive from their competitors and less prone to interference during maintenance, it is less costly to maintain them than to maintain their simpler counterparts, as supported by the presence of the sustained negativity.

However, there is one important caveat for this interpretation. Unlike previous reports of the SAN (Cruz Heredia et al., 2022; Fiebach et al., 2002; King and Kutas, 1995; Phillips et al., 2005), the topographic distribution is more global and not exclusively anterior. A possible alternative explanation for the global sustained negativity is the artifact from re-baselining (Steinhauer and Drury, 2012). Since the preceding materials for the complex and simple NP conditions are very different (two pre-modifiers in the complex NP conditions, *It appears that* in the simple NP conditions), this lexical difference could account for the sustained signal after re-baselining, as discussed by Steinhauer and Drury (2012). We acknowledge this weakness of our stimuli design, which we will keep in mind for subsequent follow-up experiments.

Nevertheless, it is important to note that previous EEG studies have also reported sustained negativities that are not anterior topographically in contexts where SAN is expected (Cruz Heredia et al., 2022; Phillips et al., 2005). As seen in Fig 3.3, Phillips et al. (2005) found that shorter filler-gap dependency lengths led to more posterior sustained negativity. Cruz Heredia et al. (2022) also detected sustained negativities in posterior electrodes for embedded *wh*-questions (relative to embedded *yes-no* questions) and for sentences with subordinating adverbs such as *after* (relative to those with non-subordinating adverbs such as *today*). For the latter, the topographic map showed a largely global negativity, except for prefrontal electrodes. They hypothesized that the different topographies reflected WM processes related to different information types: anterior negativities reflected non-linguistic referential processing while posterior negativities reflected syntactic processing. Since engagement with our stimuli required participants to maintain the NP both as a syntactic unit for further sentence building and as a referent for meaning interpretation (as evident in the behavioral experiments outlined in Chapter 2), these parallel processes can explain the presence of both anterior and posterior negativities, giving rise to a sustained negativity that was more global. However, it is unclear why other dependencies that engaged with both referential and syntactic processing only showed anterior or posterior negativities but not both (with the exception of sentences with subordinating adverbs as investigated by Cruz Heredia et al. (2022)). More EEG examination with different types of long-distance dependencies is needed to better understand the functional significance of these sustained signals.

When comparing the results in Epoch 1 with the corresponding region in Experiment 1a, the interaction between filler NP's complexity and RC type was notably missing. In Experiment 3, both sentences with SRC and ORC showed robust sustained global negativities while in Experiment 1a, despite the main effect of NP complexity on the preverbal adverb in the maintenance period, the facilitation effect was more pronounced in ORC than SRC. Based on this discrepancy between SRCs and ORCs in Experiment 1a and supporting results from Experiments 1b, 2a and 2b, we hypothesized that the aforementioned discrepancy was due sentences with SRCs have no competing referents at the preverbal adverb while sentences with ORCs have one: the RC subject NP, prompting the conclusion that representational complexity increases distinctiveness between referents and thus helps reduce interference between referents during maintenance.

Nevertheless, why was that RC type difference absent in the sustained negativity? While there was a difference in the magnitude of the facilitation effect in Experiment 1a, it is important to recall that both sentences with SRC and ORC still showed a facilitation effect when the filler NP is complex. We argued that the presence of the facilitation effect even in sentences with SRC, which do not have competing referents at the preverbal adverb, indicates that higher representational complexity also aids WM maintenance in other ways besides increasing distinctiveness between referents, such as increasing salience and/or attention due to more encoding effort or time (Karimi et al., 2020). Since attention and WM are tightly connected (Oberauer, 2019), the attention- or salience-based benefit has an effect on WM maintenance cost from the onset of the maintenance period, thus changing the magnitude of the sustained signal for both sentences with SRCs and ORCs. The benefit of increasing distinctiveness did not take place until the competing referent is encoded into WM in sentences with ORC at the subject RC NP *Andy* and thus did not generate an interaction at Epoch 1.

Epoch 2, which mostly overlapped with Epoch 1 but had the same words pre-baseline in all conditions, was analyzed in view of the pre-baselining issue with Epoch 1. We observed a sustained left anterior positivity in the simple filler NP conditions relative to the complex filler NP conditions that gradually decreased in magnitude. This ran counter to our prediction of the presence of a sustained negativity in the simple filler NP conditions, given that the Experiments 1a, 1b, 2a and 2b all showed that having a complex filler NP can facilitate the maintenance process. I speculate that this result was actually consistent with our prediction. If a sustained anterior negativity was already present at the word *warden* in the simple filler NP conditions, pre-baselining based on this word could have resulted in the opposite pattern in the EEG signal, explaining the sustained positivity with similar topography to the SAN. Thus, this analysis could lend some support (albeit limited) for the existence of the SAN, which demonstrated that maintaining complex filler NPs is less costly than maintaining simple ones.

In Epoch 3, which examined the retrieval sites for the filler NP (the RC verb *reprimanded* and the matrix verb *admitted*), we surprisingly found no evidence of a P600 or any difference between the sentences with complex filler NPs and those with simple filler NPs, indexing that the EEG results did not detect any retrieval facilitation when the filler NP is complex. While this retrieval site facilitation was found in previous related behavioral experiments (Hofmeister, 2011; Hofmeister and Vasishth, 2014; Karimi et al., 2020,1), our results from four behavioral experiments in Chapter 2 were mixed. Experiments 2a and 2b, which used coordinated filler NPs as complex NPs, did not show any retrieval site effect while Experiments 1a and 1b, which used pre-modified filler NPs as complex NPs, did find retrieval facilitation when the filler NP is complex. The inconsistency of the retrieval facilitation effect might have led to the lack of a difference in Epoch 3.

3.3.2 On Relative Clause Type

The comparison between sentences with ORC and SRC revealed 2 signals at Epoch 1 (a transient negativity on *today/Andy* at midline/left electrodes in ORC conditions and a following sustained frontal positivity) and 2 signals at Epoch 3 (a transient positivity on *Andy/today* at centro-parietal electrodes and a following sustained left frontal negativity). The two transient signals most likely resulted from the lexical differences in that time window since they occurred at the two windows where the words differed between the ORC and SRC conditions (proper noun *Andy* in one condition and adverb of time *today* in the other). Another supporting evidence for this source of the transient signals is that they were opposite in polarity, showing that the adverb of time *today* always led to a transient negativity relative to the proper noun *Andy*.

The more significant results from this comparison were the two sustained signals: the sustained positivity in the maintenance period and the sustained negativity in the retrieval period. Both of these sustained signals were elicited in anterior electrodes, which is the same topographical distribution as the SAN. Nevertheless, they were distinct from the SAN in different ways. Based on previous results that reported the SAN (Cruz Heredia et al., 2022; Fiebach et al., 2002; King and Kutas, 1995; Phillips et al., 2005), we would expect a sustained negativity during the ORC, which is more memory-intensive than SRC, relative to during the SRC. King and Kutas (1995) demonstrated the SAN while also using English relative clauses (Fig 3.1). To the contrary, our results in Epoch 1 showed a sustained anterior positivity in the ORC conditions relative to the SRC conditions, or a SAN in the SRC conditions relative to the ORC conditions. This suggested that the maintenance cost was higher in SRCs than in ORCs, if SAN indexes processing resources used during WM maintenance. A possible reason for this is the presence of the back-to-back adverbs *today harshly* in SRCs but not in ORCs. Some participants pointed out that such word sequence is unusual, especially at the beginning of a relative clause. That infrequent sequence of words might have led to additional strain to WM, which outweighed the difficulty of reading an ORC and was reflected as the sustained negativity. It is also possible that the SAN in SRCs here reflected a non-WM process entirely, implicit prosody, as proposed by Lau (2018) following the works of Steinhauer and Friederici (2001)and Steinhauer (2003). The fronting of adverb of time is often accompanied by prosodic stress on it, which could have manifested as the SAN during reading. Whatever the functional interpretation of the SAN is, the results in Epoch 1 showed the opposite pattern from King and Kutas (1995), the only work who has shown the SAN in English ORCs relative to SRCs. Our stimuli diverged from theirs in one important way: we tried to keep words in the same 500-ms time window (except for *today* and *Andy*) aligned to minimize lexical confounds while King and Kutas (1995) had different words in the same location in the two RC types. It is unclear how much lexical confounds contributed to the SAN observed by King and Kutas (1995). Further research is needed to examine the sustained negativity (or positivity) in English ORCs without lexical confounds or back-to-back adverbs.

However, our results converged with those from King and Kutas (1995) in Epoch 3, during and after matrix verb *admitted*, where the filler NP is retrieved. We also found a sustained negativity in anterior electrodes at this region in sentences with ORCs. King and Kutas (1995) observed that the SAN in ORCs persisted past the matrix verb past the end of the epoch they examined. However, Fiebach et al. (2002) and Phillips et al. (2005) did not observe SAN past the retrieval site. It is unclear what this sustained negativity after retrieval is indexing, since the filler no longer needs to be maintained after its retrieval at the matrix verb. One possible explanation is that readers continue to hold the filler in WM in case it is needed further (such as in an ellipsis). This reason is not convincing since many studies did not observe the SAN past the retrieval site. It is also possible that the sustained signal reflects the spillover effect from the lexical differences at *Andy*/*today*. Notably, not only were the ERP signals on *Andy*/*today* and *today*/*Andy* opposite in directionality, the following sustained signals were also opposite in directionality: a sustained negativity in ORCs after *Andy*/*today* and a sustained positivity after *today*/*Andy*. While this is probable, we note that the topographical distributions on *Andy*/*today* and *today*/*Andy* were not consistent with each other but the distributions of the sustained signals were (both being anterior signals). Lastly, the SAN here can also index particular retrieval processes that we do not yet understand. Tung and Brennan (2023) also found a sustained negativity at the retrieval site when the target referent and the competing referent matched in animacy, thus generating more retrieval interference than when they mismatched in animacy. The absence of a P600 effect for ORCs in both our experiment and King and Kutas (1995) might point to the SAN at retrieval as indexing retrieval difficulty instead.

3.4 Conclusion

In a follow-up EEG experiment to the behavioral experiments in chapter 2, we investigated how the encoding, maintenance and retrieval of complex filler NP differ from those of simple filler NP through an ERP lens, with the additional goal of examining how these effects interact with the type of relative clause. Our experiment revealed a sustained global negativity during the maintenance period in conditions with simple NP, relative to conditions with complex NP, showing that information about representational complexity of the filler NP is actively maintained in WM. Additionally, our comparison between ORCs and SRCs revealed a SAN for SRCs during the maintenance period and a SAN for ORCs after retrieval at the matrix verb. These results are inconsistent with previous observations of the SAN, thus warranting more experimentation to explore the nature of SAN in particular and sustained signals in general and their occurrence in relative clauses.

CHAPTER 4
THE EFFECT OF INTERNAL COHERENCE ON WORKING MEMORY PROCESSES

4.1 Introduction

In Chapters 2 and 3, we have examined how the complexity of linguistic representations, particularly that of matrix subject noun phrases, influences the three WM processes: encoding, maintenance and retrieval. By and large, more complex NPs facilitate WM maintenance and to a lesser extent, WM retrieval but have variable effects on WM encoding (depending on the type of complex NP). We have shown that having a more complex NP, which has more semantic/syntactic features, can increase distinctiveness between the NP and other competing referents and reduce interference between the referents, thus benefiting the WM maintenance and retrieval processes. However, is having more features always better for maintenance and retrieval during sentence comprehension? There exists instances where the features making up the NP do not usually occur together or are even at odds with each other. One such extreme example is *the animal-loving butcher*, where being animal-loving and being a butcher can be considered as incompatible. In this chapter, through two SPR and one EEG experiments, we examined how the compatibility between features within the complex NP, which we termed **internal coherence**, influences the costs of the three WM processes.

The following subsection 4.1.1 outlined previous research about the effects of internal coherence and the related property of similarity on WM processes in the verbal and visual domain. It aims to give the readers the backdrop to compare and contrast with the effects observed in sentence comprehension, which will be detailed in subsection 4.1.2. Subsection 4.1.2 will specifically highlight the distinction between the increased WM effort when features are shared between competing representations and the decreased WM effort when features are coherent within a representation, the latter being the phenomenon studied in this chapter.

4.1.1 The Effect of Internal Coherence and Similarity on Verbal and Visual Working Memory

Similarity is a related concept to coherence since they both have to with relatedness of features. Two objects are similar when they share certain features. For example, *cat* and *mat* are phonologically more similar than *cat* and *dog* because they share the sounds [æ] and [t] while *cat* and *dog* do not share any sounds. However, *cat* and *dog* are semantically more related because they are both pets. While internal coherence, in our definition, relates to how compatible features within a representation are, similarity relates to how features accross are related to each other. Due to this tight link between the two concepts, this section will examine how the various WM literature (visual, verbal and sentence comprehension) have investigated both coherence and similarity.

The earliest examination of similarity within WM memorandum came from the verbal WM literature, with two signature effects: the phonological similarity and the semantic similarity effect. The phonological similarity effect refers to the decrease in performance during serial recall when the items (letters or words) in the memorandum are phonologically similar relative to when they are not. For example, Conrad (1964) found that the recall accuracy for memorandum consisting of similar-sounding letters such as *b*, *p* and *v* was lower than that consisting of different-sounding letters such as *r*, *c* and *m*. Baddeley (1966) observed a similar negative effect when the memorandum made up of phonologically similar words such as *cat*, *fad* and *map*. However, the opposite effect was observed if the words in the memorandum rhymed (*cat*, *mat* and *hat*) (Fallon et al., 1999; Gupta et al., 2005). Gupta et al. (2005) hypothesized that rhyming provides a categorical cue which aids retrieval from WM while non-rhyming phonologically similar words cause interference between items during articulatory rehearsal (Hanley and Bakopoulou, 2003).

Contrary to the phonological similarity effect, the semantic similarity effect refers to the facilitation effect during recall when the memorandum consists of words that are semantically related to each other (Huttenlocher and Newcombe, 1976). Similar to rhyming words, the semantic similarity could have provided a categorical cue that aids the retrieval process.

However, a meta-regression study by Ishiguro and Saito (2021) showed that the facilitatory semantic similarity effect observed in previous works was caused by confounding definitions of semantic similarity. Previous studies were confounding between semantic similarity (the number of shared semantic features) and semantic association (the degree of associative relatedness or whether they are from the same taxonomic category). The study showed that semantic similarity imposed a negative effect on order memory (where the words are in a sequence) while semantic association led to a positive effect on item memory (what words are in the memorandum). It is also important to note that the bulk of work on similarity in verbal WM relied on the reproduction of the serial order of the memorandum.

Item similarity also produced mixed effects on WM retrieval in the visual WM domain. Lin and Luck (2009) used change detection tasks to test whether similarity in color between items in the memorandum aids or harms performance. Participants viewed a sample array of three or four colored squares. After a delay period, they were shown a test array that was either identical to the sample array or had one square of a different color. There were two types of sample arrays: homogeneous and heterogeneous. The colors of the squares in a homogeneous array were picked from different shades of the same color (red, green or blue). The colors of the squares in a heterogeneous array were picked from shades of different colors. The study showed that seeing the homogeneous array improved participants' performance. Two subsequent experiments showed that this effect was consistent even when the squares were presented sequentially or when one of the squares had a different color in the homogeneous condition. Given these results, the authors questioned the role of interference between similar items and proposed different hypotheses for the facilitatory effect of similarity. Firstly, the maintenance of similar colors could have inhibited regions between color representations and produced sharpened representations for the memorandum items. Secondly, the maintenance of multiple shades of one color could have provided an anchoring effect to prevent drifting of memory representations to the prototypical value of that color. Thirdly, memorizing similar colors could have helped better focus attention to a smaller color space, improving WM performance.

Mate and Baqués (2009) also found a facilitatory effect of similarity when the items in the memorandum are more similar. Instead of the change detection task, they presented participants with memorandum of 2,3,4 or 5 Chinese characters and asked them to identify the characters they saw from a selection of six alternatives after a delay period. The results demonstrated that similarity between the characters in the sample memorandum improved performance while similarity between the characters at test impaired performance. Additionally, they discovered that the facilitation effect from similar items in the sample memorandum increased with the set size. They hypothesized that the similarity between items in the sample memorandum allowed the processing of the items along more common dimensions and thus, lower the WM load by “simplifying the global representation" of the items. Brady and Tenenbaum (2013) similarly asserted that similarity between items helps facilitate grouping during encoding. Other works in the visuospatial domain have reported a similar grouping benefit (which they termed *perceptual grouping*) from both similarity and coherence between items in a memorandum (Brady and Tenenbaum, 2013; Diaz et al., 2021; Gao et al., 2016; Jiang et al., 2004; Mate and Baqués, 2009; Woodman et al., 2003). For example, Diaz et al. (2021) used a change detection task where the memorandum consisted of 2 or 4 circles with rectangular gaps. The additional manipulation was that in one condition, the rectangular gaps were facing each other while in another, they were misaligned. Behaviorally, they observed that the aligned trials showed higher accuracy than misaligned ones and the facilitation effect was larger for set size 4 than set size 2. Neurally, they showed that parieto-occipital electrodes displayed more negativity for 300 ms when the gaps were misaligned. The authors thus demonstrated that having objects that are more compatible for grouping (and in our words, more coherent) can help lower the WM load during both during maintenance (less negativity for the negative slow wave) and retrieval (higher response accuracy).

Yet, similar to the verbal domain, there has also been evidence for an inhibitory similarity effect in visual WM. In a verbal serial recall task, Logie et al. (2000) found that memorizing phonologically and orthographically similar words (*fly, cry, ply, dry, try, shy*) resulted in lower recall performance than memorizing phonologically similar but orthographically less similar words (*guy,thai,sigh,lie,pi,rye*). Avons (1999); Saito et al. (2008); Smyth et al. (2005)

also observed similar inhibitory similarity effects when investigating sequences of matrix patterns, Japanese kanji characters and faces, respectively. Thus, experiments in visual WM seemed to indicate conflicting effects of item similarity. However, Mate and Baqués (2009) proposed that this discrepancy can be explain by the nature of the task. The experiments that showed inhibitory effect of item similarity (Avons, 1999; Logie et al., 2000; Saito et al., 2008; Smyth et al., 2005) required the retrieval of the serial order of the items while the experiments that showed facilitatory effect did not (change detection task in Lin and Luck (2009) and recognition task in Mate and Baqués (2009)). When serial order is involved, items has to be retrieved in order and thus, their similarity causes their representations to interfere with each other, just like when the items at test are dissimilar (Mate and Baqués, 2009). This important distinction approximates our own discrimination between coherence and similarity. Internal coherence is likely to benefit WM processes while similarity is likely to harm WM processes if the task requires the retrieval of only a subset of the items. As we will see in the subsequent subsection, this distinction was also demonstrated in the sentence comprehension literature.

4.1.2 The Effect of Internal Coherence and Similarity on Working Memory During Sentence Comprehension

In the WM in sentence comprehension literature, similarity is often discussed in terms of similarity-based interference during retrieval. One of the major models exploring the interplay of WM and sentence comprehension, the Lewis and Vasishth model (hereafter abbreviated as the LV05 model) (Lewis and Vasishth, 2005), relies on feature similarity to explain many psycholinguistic phenomena. The model, which was adapted from the content-addressable ACT-R architecture, posited that syntactic chunks or constituents are encoded as bundles of features into WM. At the retrieval site, a subset of these features are used to match with retrieval cues in order to retrieve the right chunk for continuation of sentence structure building. When there are other chunks that share features with the target chunk or in other words, that are similar to the target chunk, retrieval cues can match or partially match multiple chunks in WM, thus resulting in interference during retrieval.

Many previous studies have reported similarity-based interference during retrieval (Arnett and Wagers, 2017; Cunnings and Sturt, 2018; Lowder and Gordon, 2014; Mertzen et al., 2023; Van Dyke, 2007). These studies examined instances of either or both syntactic and semantic interference - interference due to competing representations sharing similar syntactic or semantic features, respectively. For example, Mertzen et al. (2023) carried out eye-tracking studies examining both semantic and syntactic interference in both English and German. A set of English sentences in the study was as follows:

(46) a. **+animate, +subject:** It turned out that **the attorney** whose secretary had forgotten that **the visitor** was important frequently **complained** about the salary at the firm.

b. **-animate, +subject:** It turned out that **the attorney** whose secretary had forgotten that **the meeting** was important frequently **complained** about the salary at the firm.

c. **+animate, -subject:** It turned out that **the attorney** whose secretary had forgotten about **the important visitor** frequently **complained** about the salary at the firm.

d. **-animate, -subject:** It turned out that **the attorney** whose secretary had forgotten about **the important meeting** frequently **complained** about the salary at the firm.

At the retrieval site, the verb *complained*, readers needed to retrieve the target NP *the attorney* as its subject in order to comprehend the sentence. The NP *the attorney* has the syntactic feature *+subject* since it is in the subject position syntactically and the semantic feature *+animate* since it is referring to an animate referent. At retrievel, the retrieval cues include both *+subject* since the verb *complained* needs a subject and *+animate* since the verb *complained* takes an animate subject most of the time. In this setup, the authors manipulated the syntactic and semantic features on an intervening NP, *the visitor/ the meeting*. In sentence (46a), *the visitor* also had both *+animate* and *+subject* features while in sentence (46d), *the meeting* had neither features since it was neither animate or in a subject position. In sentences (46b) and (46c), the intervening NP only had *+subject* or *+animate*, respectively. Thus, the intervening NP was both semantically and syntactically similar to the target NP in (46a), only semantically similar in (46c), only syntactically similar in (46b) and neither semantically nor syntactically similar in (46d). If similarity-based interference at retrieval occurs for both syntactic and semantic features, we would expect the most reading difficulty when the intervening NP is most similar to the target NP, sentence (46a), then (46b) and (46c), and the least reading difficulty in (46d), when the intervening NP is least similar.

Indeed, in English, the authors found both reading slowdown and increased eye movement regressions at the retrieval site, the verb *complained*, both when the intervening NP is *+subject* and *+animate*. Additionally, it is relevant for our purpose to note that at the pre-retrieval site, the adverb *frequently*, they also found exact same effects, reading slowdown and increased regressions both when the intervening NP is syntactically and semantically similar to the target NP. This mirrors the site, also preverbal adverb, that we observed "maintenance" effects in the SPR experiments in Chapter 2. The results in German largely showed the same pattern in the pre-retrieval site but the retrieval site only showed effects of syntactic interference while effects of semantic interference were delayed to the post-retrieval site. The distinction between the timeline of syntactic and semantic interference is, albeit interesting, less relevant for our purpose so we will focus more on the general pattern of interference itself. In both English and German, Mertzen et al. (2023) found similarity-based interference in both the maintenance (or pre-retrieval) and retrieval sites. Similarity-based interference, as observed in sentence comprehension, resembles the inhibitory similarity effects in serial recall tasks described in Section 4.1.1, where there are competition between representations.

There are currently two groups of theories trying to explain interference effects in sentence comprehension, namely retrieval interference and encoding interference. The model of Lewis and Vasishth (2005) proposed retrieval interference as an explanation, where the existence of similar representations in WM causes cue overload (Nairne, 2002) and results in a higher probability of the comprehender retrieving the wrong representation, resulting in lower accuracy in offline measures and slower reading time at retrieval in online measures (since comprehenders need to reanalyzed upon mis-retrieval). On the other hand, proponents of encoding interference (Gordon et al., 2001) rely on the idea of feature overwriting (Nairne, 1990, 2002; Oberauer and Kliegl, 2006). During encoding, representations that share features compete over them, resulting in the degradation of the "losing" representation or both representations, leading to lower activation level and thus, difficulty during retrieval. Encoding and retrieval interference are not mutually exclusive and some previous works have shown that they might work in tandem (Mertzen et al., 2023; Villata et al., 2018). In fact, encoding interference (and retrieval interference) was used to explain the pre-retrieval reading difficulty observed by Mertzen et al. (2023). They argued that this could be the result of a spillover from the difficulty of encoding similar representations since the representations are degraded. Alternatively but not mutually exclusively, it could reflect retrieval interference due to predictive processing, the top-down process where the comprehender retrieves the target representation preemptively because they are predicting an upcoming retrieval site. The comprehender could have tried to retrieve the target NP at the the preverbal adverb, resulting in similar interference effects as the retrieval site. Mertzen et al. (2023) argued that both these explanations are possible causes for the pre-retrieval interference effects and in fact, might even work together. For our purposes, encoding interference is especially important since many effects we are interested in happened before the retrieval site.

While similarity-based interference has been relatively well-studied in the sentence comprehension literature, the effect of internal coherence within a representation to WM processes is less studied. The only experiment that has investigated this interaction is Experiment 3 from Hofmeister (2011). In Experiment 1, Hofmeister (2011) examined the effect of representational complexity by comparing between sentences with pre-modified NPs and un-pre-modified NPs. In Experiment 3, he questioned whether pre-modified NPs with atypical modifiers, such as *the lovable military dictator*, still lead to facilitation during retrieval compared to un-pre-modified NPs. The sentences from one set were as follows:

(47) a. **Simple:** The diplomat contacted **the dictator** who the activist looking for more

contributions **encouraged** to preserve natural habitats and resources.

b. **Complex, typical:** The diplomat contacted **the violent military dictator** who the activist looking for more contributions **encouraged** to preserve natural habitats and resources.

c. **Complex, atypical:** The diplomat contacted **the lovable military dictator** who the activist looking for more contributions **encouraged** to preserve natural habitats and resources.

At encoding, the head noun *dictator* and the two spillover words *who the*, average RT was slowest for sentence (47c), then (47b) and was fastest when the NP was not pre-modified in (47a). However, at the retrieval site, only (47b) showed a facilitation effect compared to (47a) while the difference between (47a) and (47c) was insignificant on the verb *encouraged* and (47a) was read marginally faster in the spillover region. These results demonstrated that more complex NPs do not always facilitate retrieval and that the coherence between features within a representation also plays a role. The experiment confirmed that representations with high internal coherence (*the violent military dictator*) facilitated both encoding and retrieval relative to representations with low internal coherence (*the lovable military dictator*).

The experiments in this chapter - Experiments 4a, 4b and 5 - examined how internal coherence and representational complexity of the target representation influence the processing effort during the three WM processes. The first goal of these experiments was to replicate and expand on the results regarding internal coherence of Experiment 3 of Hofmeister (2011). While that experiment demonstrated increased effort to encode and retrieve an incoherent representation, not much is known about how maintaining an incoherent representation affects processing cost. Since studies in other domains have shown that grouping together similar items can be beneficial for the maintenance process (Diaz et al., 2021), it is possible that grouping together coherent or similar features into a single linguistic representation can also lighten WM load and facilitate the maintenance process. Examining the effect of internal coherence of representations on encoding, maintenance and retrieval allows us to better understand how linguistic representations are grouped together in WM and to further decode WM architecture during sentence comprehension.

The second goal of these experiments was to juxtapose the effects of feature similarity/coherence between competing representations and within a representation. While much work has shown similarity-based interference when competing linguistic representations share features (Lewis and Vasishth, 2005), it is unclear whether representations that are grouped together and retrieved as a unit experience similarity-based interference. Mate and Baqués (2009) argued that for visual WM, similarity between competing representations worsens retrieval while similarity between representations within an ensemble improves it. It is possible that this distinction also extends to WM processes involving linguistic representations. In our experiments, we manipulated the degree of similarity between the referents that were grouped together in a coordinated NP. Observing how having coordinated NPs made up of similar or dissimilar referents affects WM processes can further inform us about how linguistic features are packaged within an ensemble.

Lastly, these experiments also further investigated how complex linguistic representations are encoded, maintained and retrieved, in a different sentential context from experiments in Chapters 2 and 3. We wanted to examine which effects observed in previous experiments could or could not be generalized to simpler sentential contexts in Experiments 4a, 4b and 5, which did not contain relative clauses. Understanding how representational complexity affects WM processes in different sentential setting and how it interacts with internal coherence can give us a better understanding of WM processes during sentence comprehension.

4.2 Experiment 4a

4.2.1 Participants

93 participants aged 18-50 (mean age = 31.4), recruited on Prolific, participated in this experiment for payment. All participants self-identified as native, monolingual English speakers who were raised in monolingual households. In addition, participants had no language related disorders or literacy difficulties. Data from 7 out of 93 participants was removed because they scored below the threshold of 85% for the comprehension questions (1 standard deviation below the mean accuracy, rounded to the nearest multiples of 5%).

4.2.2 Methods and Materials

The self-paced reading experiment had a 3 x 2 design. It consisted of 36 6-condition items which were manipulated in terms of the number of referents in the filler NP (3 referents vs. 2 referents vs. 1 referent) and the similarity between the referents (similar referents vs. dissimilar referents). An example sentence set is given in (48a)-(48f), with the slashes indicating the self-paced reading regions.

All sentences consisted of the matrix subject filler NP, which was followed by the parenthetical (*according to* and a human name) and then the matrix VP, consisting of *had been*, the participle form of the main verb and then a *by*-PP. The filler NP was either coordinated (2- or 3-referent conditions) or not (1-referent conditions) but the referents were all singular definite NP. In the 2- or 3-referent conditions, all the referents were either chosen from the same predetermined conceptual category (similar-referent conditions) or from different categories (dissimilar-referent conditions). The categories included were non-human animals, plants, food, locations, clothes, modes of transportation and objects. For example, in sentence (48a), *hamster*, *cat* and *dog* all came from the category of non-human animals. In contrast, in sentence (48d), *car* came from the category of modes of transportation, *TV* came from the category of objects and *dog* came from the category of non-human animals. Trivially, the similar and dissimilar 1-referent conditions were the same. The final referent in the NP was always the same referent (*the dog*) across different conditions in the same sentence set. Example sentences for Experiment 4a are as follows:

(48) a. **Similar, 3-referent:** The hamster, / the cat / and / the dog, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

b. **Similar, 2-referent:** The cat / and / the dog, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

c. **Similar, 1-referent:** The dog, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

d. **Dissimilar, 3-referent:** The car, / the TV / and / the dog, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

e. **Dissimilar, 2-referent:** The TV / and / the dog, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

f. **Dissimilar, 1-referent:** The dog, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

Each participant read only one condition per item, totaling 36 experimental sentences. In addition, they also read 36 filler sentences, which were sentences that had similar structures as the experimental sentences. The referents in the filler sentences were always from the same category and the matrix VP was in the past progressive. One such filler sentence is *The antelope, / the goat / and / the sheep, / according to / Nicole, / had / been / climbing / the hill / by / the pasture.* Each sentence, experimental or filler, preceded a yes-no comprehension question mentioning a referent. The referent mentioned could come from any of the referents in the filler NP or could be a new referent. To ensure that participants engaged in non-shallow processing, the verb was also sometimes changed in the question. If the question had either an unseen referent or an unseen verb, the answer was expected to be *No*. The expected answer for half of the questions was *Yes* and for the other half was *No*. An example question for sentences (48a)-(48f) is *Had the dog been abandoned?* (Expected answer: *Yes*). Participants did not receive feedback on the accuracy of their answers. In this experiment, average accuracy across all items (including fillers) was 93.1% , and the average accuracy on the experimental sentences was 93.7%.

The experiment was carried out on Ibex Farm (Drummond, 2013). Participants did two practice trials before reading the experimental and filler items in a randomized order. Before a trial started, a dash line appeared in the middle of the screen where the stimuli would appear. Upon pressing the space bar, the dash line disappeared and the first word appeared. Participants were instructed to press the space bar to continue reading the sentence. As the space bar was pressed, the current word(s) was replaced by the subsequent word(s).

For statistical analysis, after excluding data from participants who did not meet the comprehension accuracy threshold of 85%, raw RT beyond three standard deviations of the mean raw RT at each sentence position and condition were excluded. We rejected 0.60% of the raw RTs through this procedure. Following Hofmeister and Vasishth (2014), we did not exclude RTs of sentences whose comprehension question was answered incorrectly. This was done to ensure we did not discard instances that might have reflected failure to maintain the correct dependency in WM.

RTs in each region were log-transformed and then residualized on two predictors: the linear position of a region in a sentence and log RT of the region immediately prior to the current one. Both predictors are known to impact self-paced reading RTs for independent reasons. Residualization was done using linear mixed models through the *lmer* package (Bates et al., 2015). RTs from fillers were included in the residualization process. Residualized log RTs served as the dependent variable for our analyses.

For data analysis, we considered three sites of interest - encoding, maintenance and retrieval sites. The **encoding site** included the coordinator *and* and the final referent of the filler NP *the dog*. We only analyzed this region for the 2- and 3-referent conditions because the 1-referent conditions did not have *and* and RT for *the dog* in those conditions could not be residualized based on the previous word's RT. RTs at this region reflected processing effort to encode the subject noun phrase into WM representation. The **retrieval site** included the VP up to the verb *had been recently abandoned*. At the retrieval site, the filler NP needs to be retrieved from WM to serve as the object of the main verb. We also analyzed the **maintenance site**, which included the words between the encoding and retrieval sites - the parenthetical *according to Mary*. During this period, the filler NP is maintained in WM, awaiting for retrieval.

For statistical analyses, we employed Bayesian hierarchical modeling, using the R package *brms* (Bürkner, 2017). For each self-paced reading region examined, the model used 4 chains, with 2000 samples per chain, the initial 1000 samples being warm-up samples and no thinning. This led to 4000 post-warmup samples for each parameter estimate per region. Models of all word regions comprised of fixed effects of number of referents, referent similarity (sum coded, similar -0.5, dissimilar 0.5) and their interaction. For *and* and *the dog*, number of referents *Number (2 vs 3)* was sum coded as 2-referent 0.5 and 3-referent -0.5. For the rest of the words, number of referents was backward difference-coded as *Number (2 vs 3)* and *Number (1 vs 2)*. All models also contained by-participant and by-item random intercept adjustments and random slopes for all fixed effects analyzed in that region. We used relatively weak, uninformative priors for all parameters. For the prior for all the fixed effects, including the intercept, we used a normal distribution with mean of 0 and standard deviation of 10[1]. The final results were reported in terms of the mean of the posterior distributions and the 95% credible intervals. We did not report the interaction of *Number (1 vs 2)* x similarity since the two 1-referent conditions were the same. We considered a predictor as reliable if the credible interval did not include 0.

4.2.3 Results

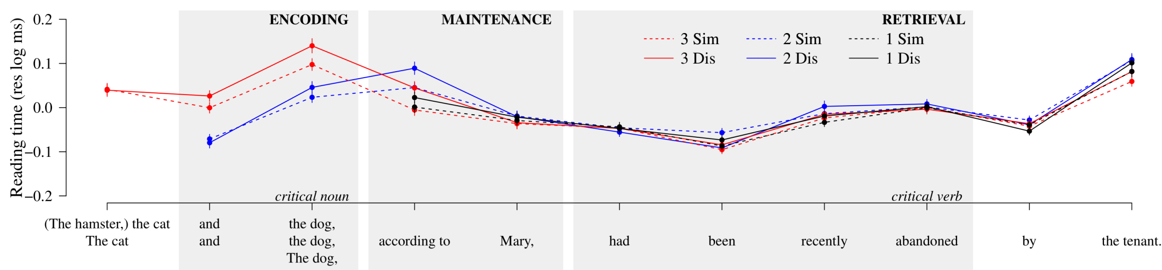

Figure 4.1: Experiment 4a: Residualized log reading times

In the **encoding site**, which only compared the 2- and 3-referent conditions, as shown in Figure 4.2 and Table 4.1, there was evidence for a main effect of referent number. Specifically, on both words *and* and *the dog*, the 2-referent “simple" conditions were read faster than the 3-referent “complex" conditions. There was weak evidence for a main effect of referent similarity on both words where the similar conditions were read faster than the dissimilar conditions.

1. An example of a *brm* model:

```
brm(formula = residualizedlogrt ~ refnumber * refsimilarity + (1 + refnumber * refsimilarity|participant)
+ (1 + refnumber * refsimilarity|item), data = rt_data, family = gaussian(),
prior = c(prior('normal(0,10)', class = 'Intercept'), set_prior('normal(0,10)', class = 'sigma'),
set_prior('normal(0,10)', class = 'b'),set_prior('normal(0,10)', class = 'sd'), set_prior('lkj(2)', class =
'cor')),
warmup = 1000, iter = 2000, chains = 4, control = list(adapt_delta = 0.99, max_treedepth = 12))
```

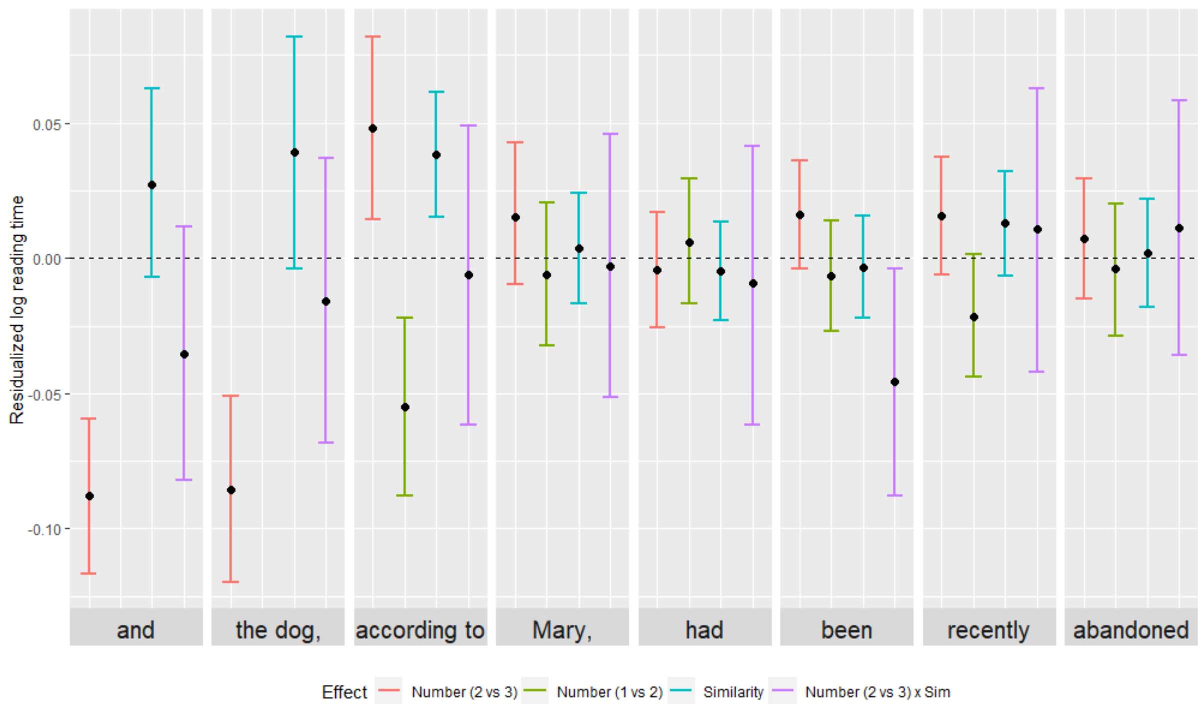

Figure 4.2: Experiment 4a: Credible intervals

In the **maintenance site**, which compared all the conditions, there was evidence for effects of both *Number (2 vs 3)* and *Number (1 vs 2)* on the word *according to*. Specifically, both the 3-referent and the 1-referent conditions were read faster than the 2-referent conditions on this word. There was evidence for a main effect of referent similarity on *according to* where participants read faster when the referents were similar. There was no effect detected on *Mary*.

In the **retrieval site**, there was no evidence for any effect except for an interaction between *Number (2 vs 3)* and similarity on the word *been*. The interaction was likely to be spurious because it involved the 2-similar-referent condition being read slower than the 2-dissimilar-referent condition on this region, a trend not reflected in any other regions. Additionally, Experiment 4b, which had a similar design to this experiment, did not replicate this interaction. *been* and *recently* both showed weak evidence of a main effect of *Number (2 vs 3)*, where the 3-referent conditions were read faster than the 2-referent conditions. On *recently*, there was also weak evidence for a main effect of *Number (1 vs 2)* (1-referent conditions were read faster than 2-referent conditions) and similarity (similar-referent conditions

Region	Effect	Mean	CrI lower	CrI upper
and	Number (2 vs 3)	-0.088	-0.117	-0.060
	Referent similarity	0.027	-0.007	0.063
	Number (2 vs 3) x Similarity int.	-0.035	-0.082	0.012
the dog,	Number (2 vs 3)	-0.086	-0.120	-0.051
	Referent similarity	0.039	-0.004	0.082
	Number (2 vs 3) x Similarity int.	-0.016	-0.068	0.037
according to	Number (2 vs 3)	0.048	0.014	0.082
	Number (1 vs 2)	-0.055	-0.088	-0.022
	Referent similarity	0.038	0.015	0.061
	Number (2 vs 3) x Similarity int.	-0.006	-0.062	0.049
Mary,	Number (2 vs 3)	0.015	-0.010	0.042
	Number (1 vs 2)	-0.006	-0.033	0.020
	Referent similarity	0.003	-0.017	0.024
	Number (2 vs 3) x Similarity int.	-0.003	-0.051	0.046
had	Number (2 vs 3)	-0.004	-0.026	0.017
	Number (1 vs 2)	0.006	-0.017	0.029
	Referent similarity	-0.005	-0.023	0.013
	Number (2 vs 3) x Similarity int.	-0.009	-0.062	0.041
been	Number (2 vs 3)	0.016	-0.004	0.036
	Number (1 vs 2)	-0.007	-0.027	0.014
	Referent similarity	-0.003	-0.022	0.016
	Number (2 vs 3) x Similarity int.	-0.046	-0.088	-0.004
recently	Number (2 vs 3)	0.016	-0.006	0.037
	Number (1 vs 2)	-0.022	-0.044	0.001
	Referent similarity	0.013	-0.006	0.032
	Number (2 vs 3) x Similarity int.	0.011	-0.042	0.063
abandoned	Number (2 vs 3)	0.007	-0.015	0.030
	Number (1 vs 2)	-0.004	-0.029	0.020
	Referent similarity	0.002	-0.018	0.022
	Number (2 vs 3) x Similarity int.	0.011	-0.036	0.058

Table 4.1: Experiment 4a: Model summary for each region and fixed effect or interaction examined

were read faster than dissimilar-referent conditions).

4.2.4 Discussion

In this experiment, we investigated how the representational complexity and internal coherence of the NP influenced the three WM processes and how they interact with each other.

Representational complexity was manipulated by varying the number of coordinated referents in the filler NP, with NP having more referents being more complex. Internal coherence was manipulated by choosing referents that come from the same conceptual category or from different categories.

Regarding representational complexity, in this experiment, we examined three groups with NPs with increasing complexity that have one, two or three referents, respectively. In the **encoding site**, sentences with 3-referent NPs were read slower than those with 2-referent NPs. This result aligned with previous results (Hofmeister, 2011; Hofmeister and Vasishth, 2014; Karimi et al., 2020) and our own results from Experiments 1a and 1b that having a more complex NP increases encoding effort. However, the fact that these results contradict with Experiments 2a and 2b, which also used coordinated NP as more complex NP but showed that having more referents sped up the RT at the encoding site. On the first region of the **maintenance site** *according to*, we found that having either 3 referents or 1 referent facilitated reading. From the experiments in Chapter 2, we would expect that increasing the number of referents, or representational complexity of the NP, would lead to less maintenance cost. However, that the 1-referent conditions were read faster than the 2-referent conditions was unexpected. In the **retrieval site**, the interaction between *Number (2 vs 3)* and similarity on the word *been* appeared to be resulted from the similar 2-referent condition being read slower than the other three conditions. Given no evidence for main effects in the retrieval site, this interaction was likely spurious. Lastly, some words in this site, *been* and *recently*, showed trends that were similar to the maintenance site, where the 2-referent conditions was read slower than the 1-referent and 3-referent conditions.

In the encoding site, we witnessed a reading slowdown with increasing number of referents, unlike Experiments 2a and 2b which detected a speed up instead. It is possible that this was due to the different degrees of lexical priming between these experiments in this site. In Experiments 2a and 2b, where the referents in the complex conditions were related to each other (such as *those judges and lawyers*), reading the first referent could have primed the second referent, making lexical access for the second referent easier and reducing reading time relative to the simple conditions. However, in this experiment, due to the referent similarity manipulation, only the referents in the similar-referent conditions were primed. Thus, the average degree of lexical priming was less than in Experiments 2a and 2b, potentially leading to the slowdown in the 3-referent conditions relative to the 2-referent conditions. However, this explanation is unlikely given the lack of interaction between number and similarity in this region. We would expect that the similar 3-referent condition showed a speed up while the dissimilar 3-referent condition showed a slowdown relative to their 2-referent counterparts but this was not the case.

An alternative explanation is that these experiments actually made different comparisons with regards to number of referents. While Experiments 2a and 2b compared between 1- and 2-referent conditions, analysis in the encoding site in Experiment 4a compared between 2- and 3-referent conditions. It is possible that 2-referent NPs were encoded faster than both 1- and 3-referent NPs, a pattern mirroring what was observed at the maintenance and retrieval sites of Experiments 4a, where 2-referent conditions were read slower than the other conditions. We could not compare with the 1-referent conditions at this site because *the dog* was sentence-initial when there was only 1 referent, a design problem that would be rectified in Experiment 4b.

A key result we observed from this experiment is the reverse trends between encoding and the later processes. The 3-referent conditions were read slower than the 2-referent conditions during encoding but were read faster during early maintenance and to a lesser extent, retrieval. This pattern is in line with the experiments in Chapter 2, which showed that the degree of facilitation during maintenance and retrieval thanks to increasing complexity of the target NP is modulated by encoding effort. These results together supported the hypothesis that increased encoding effort of the target NP can facilitate its maintenance and retrieval (Hofmeister, 2011; Karimi et al., 2020).

Regarding internal coherence of the NP, the results from Experiment 4a demonstrated that higher internal coherence within the NP decreased the effort needed for all three WM processes, especially for encoding and early maintenance. Thus, grouping referents that shared features into a representation decreased effort to encode and maintain the representation, which is in line with the results of Experiment 3 from Hofmeister (2011).

4.3 Experiment 4b

A significant drawback of Experiment 4a was that the target NP was sentence-initial. This prevented us from residualizing RTs in the encoding site *the dog* in the 1-referent conditions and from comparing the encoding site in those conditions with the 2- and 3-referent conditions. Experiment 4b sought to improve on that shortcoming by inserting additional lexical materials prior to the encoding site, thus allowing us to compare the encoding sites of all conditions. Another goal of Experiment 4b was to replicate the effects in the encoding and early maintenance sites that we had observed in Experiment 4a.

4.3.1 Participants

101 participants aged 18-50 (mean age = 31.8), recruited on Prolific, participated in this experiment for payment. All participants self-identified as native, monolingual English speakers who were raised in monolingual households. In addition, participants had no language related disorders or literacy difficulties. Data from 7 out of 101 participants was removed because they scored below the threshold of 85% for the comprehension questions (1 standard deviation below the mean accuracy, rounded to the nearest multiples of 5%).

4.3.2 Methods and Materials

The experiment had the same setup and materials as Experiment 4a, except for the fact that the filler NP was placed in an *it*-cleft to prevent it from being sentence-initial. Specifically, the filler NP in Experiment 4a was now preceded by *It was* in all conditions and the complementizer *that* was added behind the final noun *the dog*. Thus, Experiment 1b had the same 3 (referent number, sum-coded for *and* and backward-difference coded for the rest of the words) x 2 (referent similarity, sum coded, similar -0.5, dissimilar 0.5) design. Example sentences for Experiment 4b are as follows:

(49) a. **Similar, 3-referent:** It was / the hamster, / the cat / and / the dog / that, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

b. **Similar, 2-referent:** It was / the cat / and / the dog / that, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

c. **Similar, 1-referent:** It was / the dog / that, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

d. **Dissimilar, 3-referent:** It was / the car, / the TV / and / the dog / that, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

e. **Dissimilar, 2-referent:** It was / the TV / and / the dog / that, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

f. **Dissimilar, 1-referent:** It was / the dog / that, / according to / Mary, / had / been / recently / abandoned / by / the tenant.

The *it*-cleft was also added to all fillers. The comprehension questions remained the same. The sentences were presented in a self-paced reading task, like in Experiment 4a. In this experiment, average accuracy from all participants for the comprehension questions, including those following filler sentences, was 92.3%, and the average accuracy on the experimental sentences was 92.9%.

All statistical analyses followed the same procedure as Experiment 4a. We removed 0.89% of RT data that were 3 standard deviations of the mean raw RT at each sentence position and condition. The **encoding site** and **retrieval site** were the same as Experiment 4a. The **maintenance site** was extended to include the additional word *that*.

4.3.3 Results

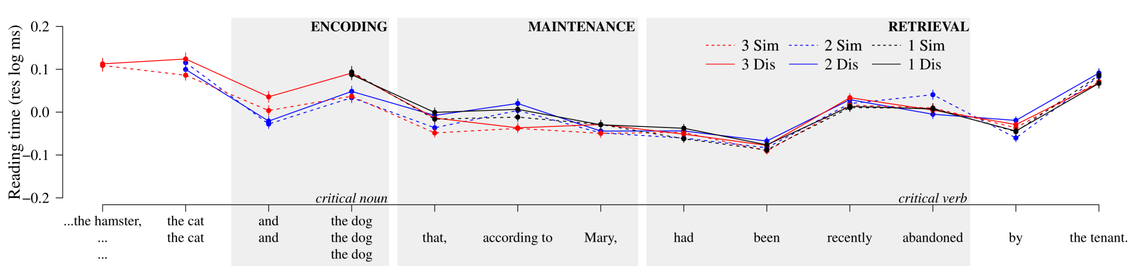

Figure 4.3: Experiment 4b: Residualized log reading times

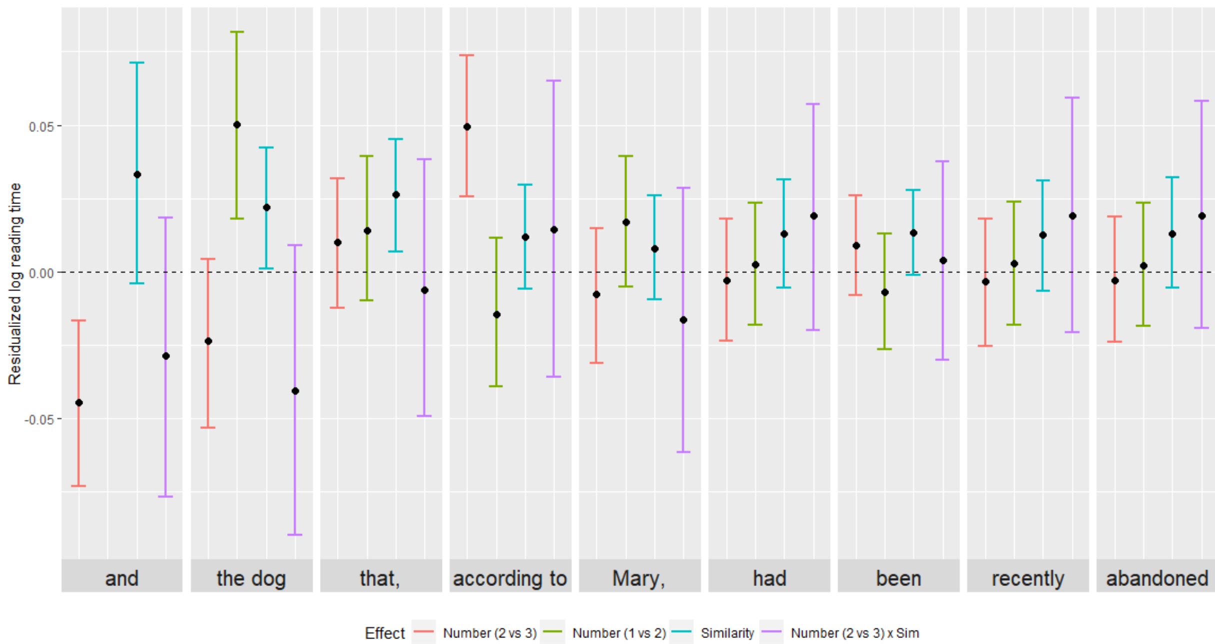

Figure 4.4: Experiment 4b: Credible intervals

In the **encoding site**, there was evidence for a main effect of referent number on *and* (where only the 2- and 3-referent conditions were compared) where the 2-referent conditions were read faster than the 3-referent conditions. On the referent *the dog*, there was weak evidence for a main effect of *Number (2 vs 3)* where the 3-referent conditions were read slower than the 2-referent conditions. Also in this region, there was evidence for a main effect of effect of *Number (1 vs 2)* where the 1-referent conditions were read slower than the 2-referent conditions. In other words, on *the dog*, the 2-referent conditions were read faster than the other conditions. For referent similarity, there was weak evidence that participants read similar-referent conditions faster on *and* and evidence that they read similar-referent conditions faster on *the dog* than dissimilar-referent conditions.

In the **maintenance site**, for referent number, there was only evidence for a main effect of *Number (2 vs 3)* on *according to*, reflecting that the 3-referent conditions were read faster than the 2-referent conditions in this region. For referent similarity, there was evidence that participants read similar-referent conditions faster on *that* and weak evidence that they read similar-referent conditions faster on *according to* and *Mary* than dissimilar-referent conditions.

Region	**Effect**	**Mean**	**CrI lower**	**CrI upper**
and	Number (2 vs 3)	-0.045	-0.073	-0.017
	Referent similarity	0.033	-0.004	0.071
	Number (2 vs 3) x Similarity int.	-0.029	-0.077	0.018
the dog	Number (2 vs 3)	-0.024	0.053	0.004
	Number (1 vs 2)	0.050	0.018	0.082
	Referent similarity	0.022	0.001	0.042
	Number (2 vs 3) x Similarity int.	-0.040	-0.090	0.009
that,	Number (2 vs 3)	0.010	-0.012	0.032
	Number (1 vs 2)	0.014	-0.010	0.039
	Referent similarity	0.026	0.007	0.045
	Number (2 vs 3) x Similarity int.	-0.006	-0.049	0.038
according to	Number (2 vs 3)	0.050	0.026	0.074
	Number (1 vs 2)	-0.015	-0.039	0.011
	Referent similarity	0.012	-0.006	0.030
	Number (2 vs 3) x Similarity int.	0.014	-0.036	0.065
Mary,	Number (2 vs 3)	-0.008	-0.031	0.015
	Number (1 vs 2)	0.017	-0.005	0.039
	Referent similarity	0.008	-0.010	0.026
	Number (2 vs 3) x Similarity int.	-0.016	-0.062	0.029
had	Number (2 vs 3)	-0.003	-0.024	0.018
	Number (1 vs 2)	0.003	-0.018	0.023
	Referent similarity	0.013	-0.006	0.031
	Number (2 vs 3) x Similarity int.	0.019	-0.020	0.057
been	Number (2 vs 3)	0.009	-0.008	0.026
	Number (1 vs 2)	-0.007	-0.026	0.013
	Referent similarity	0.014	-0.001	0.028
	Number (2 vs 3) x Similarity int.	0.004	-0.030	0.037
recently	Number (2 vs 3)	-0.003	-0.025	0.018
	Number (1 vs 2)	0.003	-0.018	0.024
	Referent similarity	0.013	-0.006	0.031
	Number (2 vs 3) x Similarity int.	0.019	-0.021	0.059
abandoned	Number (2 vs 3)	-0.003	-0.024	0.019
	Number (1 vs 2)	0.002	-0.019	0.024
	Referent similarity	0.013	-0.005	0.032
	Number (2 vs 3) x Similarity int.	0.019	-0.019	0.058

Table 4.2: Experiment 4b: Model summary for each region and fixed effect or interaction examined

Similar to Experiment 4a, in the **retrieval site**, there was no evidence for main effects of either referent number or similarity. However, there was weak evidence that similar-referent conditions were read faster than dissimilar-referent conditions for all four words in this site.

4.3.4 Discussion

Experiment 4b was done with two main objectives. Firstly, we wanted to examine whether the effects (or lack of effects) during maintenance and retrieval sites in Experiment 4a replicated. Secondly, we wanted to observe the full range of conditions, including the 1-referent conditions, in the encoding site. This was not done in Experiment 4a because the encoding site was sentence-initial in the 1-referent conditions. Thus, in Experiment 4b, an *it*-cleft was inserted at the beginning of the sentence so that we could compare the 1-referent conditions to the other conditions.

Regarding representational complexity, in this experiment, we examined all conditions with NP comprising of 1, 2 and 3 referents in all three sites. In the **encoding site**, at the word *and*, sentences with 3-referent filler NPs were read slower than sentences with 2-referent filler NPs, an effect consistent with that observed in Experiment 4a. At the region *the dog*, we observed the same trend, although the evidence was weak. Additionally, at this region, sentences with 1-referent NPs were also read slower than sentences with 2-referent NPs, which were consistent with the effect observed in the encoding site of Experiments 2a and 2b. In the **maintenance site**, at the word *according to*, participants read 3-referent conditions faster than 2-referent conditions, again consistent with the same effect at this region in Experiment 4a. However, the main effect of *Number (1 vs 2)* was absent in Experiment 4b but present in Experiment 4a, where the 1-referent conditions were read faster than the 2-referent conditions. Lastly, similar to Experiment 4a, we detected no main effects of referent number in the **retrieval site**.

During encoding, the hypothesis proposed in Section 4.2.4 that 2-referent conditions were always read slower than 1-referent and 3-referent conditions on *the dog* was confirmed. While 3-referent condition was encoded slower because of its higher representational complexity (Hofmeister, 2011), the question is why participants read the 1-referent conditions slower than the 2-referent conditions. It is possible that in the 1-referent conditions, *the dog* was the first time the participants encountered the filler NP. Thus, participants needed to start the structure building of the NP. Conversely, at this word in the 2- and 3-referent conditions, participants had already started building the structure of the filler NP earlier on and instead had to incorporate the new referent into the existing NP structure. Thus, the processing cost needed to establish the NP structure at *the dog* in the 1-referent conditions could have outweighed the cost needed to incorporate additional referents in the 2- and 3-referent conditions, leading to slower RT at this word for the 1-referent conditions. However, if this is the case, we would expect to see a similar effect between the 2-referent and 3-referent conditions at *the cat/the TV*. In other words, we would predict that at the first referent of the 2-referent conditions and the second referent of the 3-referent conditions, RT for the 2-referent conditions would be slower than the 3-referent conditions due to the additional cost of starting NP structure building. However, a post-hoc analysis showed no main effect of referent number at this region ($\beta = 0.003$, CI [-0.035, 0.040]). It remains unclear why the 2-referent conditions were read the fastest at this word. Further experimentation is needed to understand this pattern.

During maintenance, specifically on *acccording to*, we again observed a facilitation effect in the 3-referent conditions relative to the 2-referent conditions. However, the facilitation effect in the 1-referent conditions observed in Experiment 4a was absent in this experiment. It appears that the maintenance of the 3-referent conditions was facilitated, possibly because of the increased representational complexity and increased encoding time. The facilitation of the 1-referent conditions was not as robust because while the 1-referent conditions were encoded slower than the 2-referent conditions at *the dog*, its overall encoding time was shorter than the 2-referent conditions (three words shorter, *the cat/the TV and*) and the NP had lower representational complexity. This was also reflected in the less robust maintenance and retrieval site effects in Experiments 2a/2b relative to Experiments 1a/1b.

Regarding referent similarity, the results mirrored those in Experiment 4a. In the **encoding site**, the similar-referent conditions were read faster than the dissimilar-referent conditions on both words *and* (with weak evidence) and *the dog*. In the early period of the **maintenance site**, on the word *that*, the similar-referent conditions were also read faster, mirroring the facilitation effect observed on *according to* in Experiment 4a, the first region of the **maintenance site**. On the rest of the words in the **maintenance site** and all the words in the **retrieval site**, there was weak evidence that RTs in the similar-referent conditions were faster, showing a more persistent effect than Experiment 4a, where there was only a weak evidence for an effect found was at the word *been*.

The results in Experiments 4a and 4b showed that higher referent similarity or, in other words, higher NP internal coherence always has a facilitatory effect on the WM processes. However, the extent to which the three WM processes is affected differed. The facilitation was most robust during encoding and early maintenance and less robust during late maintenance and retrieval (only Experiment 4b showed weak evidence for effects during retrieval). Additionally, the facilitation during early maintenance always happened on the region immediately after the encoding site, which might suggest that the main effect might be a spillover from the encoding site. It is possible that participants were still attempting to encode the NP into WM at the first region of the maintenance site.

In order to further explore whether this facilitatory effect during early maintenance was due to reduced maintenance effort or due to a spillover from reduced encoding effort, we followed up on Experiments 4a and 4b with an EEG experiment, Experiment 5. An advantage of EEG over SPR in investigating the maintenance site is the existence of the SAN, a relatively well studied sustained ERP signal that arguably indexes processing effort during maintenance (Fiebach et al., 2002; King and Kutas, 1995; Phillips et al., 2005). If the facilitatory effect thanks to increased internal coherence during maintenance is due to a spillover from encoding, we would expect to not observe a SAN between the conditions with an incoherent target NP and the conditions with a coherent target NP. On the other hand, if increased internal coherence of the target NP decreased maintenance effort, we would predict the presence of the SAN during early maintenance, the region we observed the facilitation in Experiments 4a and 4b.

4.4 Experiment 5

Experiment 5 was an EEG experiment that followed up on Experiments 4a and 4b to take advantage of the understanding of the SAN as an ERP index of maintenance effort. It, thus, had similar design and stimuli to Experiment 4b. Additionally, Experiment 5 gave us an EEG perspective on how internal coherence and representational complexity of the target NP influence the three WM processes, allowing us to have a more multi-dimensional view to understand these processes.

4.4.1 Participants

52 participants (27, 20 and 5 identified as female, male and non-binary, respectively; 47 and 5 self-identified as right-handed and left-handed, respectively; age ranging between 18-28 years old, mean age = 20.04 years, standard deviation of age = 2.32 years) participated in this experiment for payment. All participants self-identified as native American or Canadian English speakers. In addition, participants had no language/attention/memory related disorders or literacy difficulties.

4.4.2 Stimuli

(50) **3-referent, similar:** It was / the hamster, / the cat / and / the dog/ that, / according to / Andy, / were / abandoned / by / the tenant.

(51) **2-referent, similar:** It was / the cat / and / the dog/ that, / according to / Andy, / were / abandoned / by / the tenant.

(52) **3-referent, dissimilar:** It was / the kayak, / the TV / and / the dog/ that, / according to / Andy, / were / abandoned / by / the tenant.

(53) **2-referent, dissimilar:** It was / the TV / and / the dog/ that, / according to / Andy, / were / abandoned / by / the tenant.

Sentences (50)-(53) are example sentences from the four conditions. The forward slashes separate the words or groups of words that is presented individually. The experiment had a 2 x 2 design, where number of referents (3-referent/2-referent) and similarity between referents (similar/dissimilar) were manipulated. The stimuli were designed to be similar in structure to those in Experiment 4b. Unlike in Experiment 4b, the filler NP can only have two or three referents. The one-referent conditions were removed to reduce experiment length. Like in Experiment 4b, to manipulate the similarity between referents, nouns were categorized into different predetermined conceptual categories (humans, non-human animals, plants, food, locations, clothes, modes of transportation and objects). For similar-referent conditions, all referents came from the same category. For dissimilar-referent conditions, all referents came from different categories. We also obtained the word embeddings for every referent using the *en_core_web_lg* pipeline from *Spacy* (Honnibal and Montani, 2017). We then calculated the cosine similarity between each pair of nouns from the same coordinated NP. For three-referent conditions, the three pair-wise similarity scores were averaged to yield the score for the condition. For sentence sets where the scores for the dissimilar-referent condition were higher than those for the corresponding similar-referent condition, the nouns were changed and similarity scores were recalculated. The mean similarity scores for the conditions are: 3-referent, similar: 0.600 (SD = 0.097); 3-referent, dissimilar: 0.152 (SD = 0.058); 2-referent, similar: 0.608 (SD = 0.124); 2-referent, dissimilar: 0.168 (SD = 0.091). The main verb was chosen so that it was plausible for all nouns in both similar- and dissimilar- conditions.

Half of the sentences had corresponding comprehension questions in the form of *Did the N1 V the N2?*, where N1 was always the final noun of the sentence *tenant*. The answers for half of the questions were expected to be *yes* and the answers for the other half were expected to be *no*. For questions whose answers were *yes*, N2 was counterbalanced between the referents at the three different positions to ensure that participants did not rely on heuristics such as 'Only remember the last conjunct' to answer the question. For some questions whose answers were *no*, N1 was a noun in the same category as one of the three referents but not one of the referents. To encourage comprehension of the sentence, some questions whose answers were *no* were also created by using the wrong verb. This was done so that participants had to understand the sentence and not only memorize the word list in the filler NP. 200 sets of experimental stimuli formed four stimuli lists, created by a Latin square design. Thus, participants only read one sentence per sentence set. Each stimuli list was used for 13 participants.

4.4.3 Procedure

Participants performed two experiments back-to-back: Experiment 3 detailed in Chapter 3 and this experiment. This experiment was always done second. The participant was assigned one of the four stimuli lists based on their participant number. Each trial consisted of: the fixation cross (1s), sentence presentation, comprehension question (100% of trials) and the *Ready?* screen. Sentences were presented using the RSVP paradigm, with each word or group of two words presented for 300 ms and an inter-stimulus interval of 200 ms. Participants were instructed not to move their eyes, face or body and not to read the sentence out loud during sentence presentation. In half of the trials, after the presentation of the last group of words of the sentence, a comprehension question appeared and participants had to press *f* if they thought the answer is *yes* and *j* if they thought the answer is *no*. For this experiment, participants had 4 s to indicate their choice. Participants were then encouraged to take a break and blink their eyes during the *Ready?* screen between trials and press the *Spacebar* to continue to the next trial when they are ready. The experimental trials (200 sentences) were randomly divided into five blocks (40 sentences per block). The experimenter provided feedback to the participant in between the blocks and encouraged them to take an extended break. The experimental trials were preceded by ten practice sentences, whose trial format was the same as that of the experimental trials.

4.4.4 Electrophysiological Recordings

EEGs were recorded from 32 electrodes from a Brain Vision actiCHamp Plus System located at Fp1/2, Fz, F3/4, F7/8, FC1/2, FC5/6, FT9/10, Cz, C3/4, T7/8, CP1/2, CP5/6, TP9/10, Pz, P3/4, P7/8, Oz, O1/2. A pair of electrodes (VEOG) was placed above and below the right eye and another pair (HEOG) was placed at the outer canthi of both eyes to monitor vertical and horizontal eye movements respectively. EEGs were referenced online to the average of all EEG electrodes. Impedances of all EEG electrodes were maintained below 5 kΩ for the entire duration of the experiment. Continuous data were digitized at a sampling rate of 1000 Hz.

4.4.5 Data Preprocessing and Analysis

Preprocessing of the raw data was done using EEGLAB (Delorme and Makeig, 2004) and ERPLAB (Lopez-Calderon and Luck, 2014), which are extensions of MATLAB. EEG data were band-pass filtered between 0.01 and 40 Hz, then re-referenced to the average of the two mastoid electrodes TP9 and TP10.

Since we are interested in the EEG signal that corresponds to the encoding, maintenance and retrieval periods, we epoched the EEG data into three epochs. Epochs 1 and 2 reflected the period where the filler NP was encoded into WM and maintained there. Epoch 1 was a 2500-ms epoch starting from the onset of the penultimate referent of the filler NP *the cat/the TV* and lasting until the offset of the blank screen after *according to*. Epoch 2 was a 2500-ms epoch starting from the onset of the final referent of the filler NP *the dog* and lasting until the offset of the blank screen after *were*. While Epoch 1 focused more on the incremental encoding of the coordinated NP into WM, Epoch 2 focused more on the maintenance of the filler NP in WM awaiting retrieval. Epoch 3, which reflected the period where the filler NP was retrieved from WM, was a 2000-ms epoch starting from the onset of the verb *were* and lasting until the end of the sentence. The data in all epochs were baselined by the 100 ms immediately before the epoch. For artifact rejection, we carried out peak-to-peak thresholding of the VEOG and HEOG electrodes (200 ms time window with 100 ms step) to exclude trials whose changes were more than 70 mV. We also removed trials with extreme voltages in EEG electrodes that were not within the range of [-75 mV, 75 mV]. Additionally, for each epoch individually, we excluded epochs from any participants whose rejection rate was more than 50%. After this procedure, data from 45 out of 52 participants were retained for Epoch 1, data from 46 participants were retained for Epoch 2 and data from 41 participants were retained for Epoch 3. We removed 28.5%, 29.6% and 41.1% of trials from Epoch 1, 2 and 3, respectively through this artifact rejection procedure. There were 41.1, 39.6 and 37.2 trials left on average per condition for Epoch 1, 2 and 3, respectively.

To detect reliable differences for both manipulations (number of referents and similarity between referents) between the ERPs, we used a repeated measures, two-tailed cluster-based permutation test based on the cluster mass statistic (Bullmore et al., 1999) using a family-wise alpha level of 0.01. This was done using the Mass Univariate ERP Toolbox (MUT) (Groppe et al., 2011), an extension of MATLAB. The data was first downsampled from 1000 Hz to 250 Hz using boxcar filter and then re-baselined using the 100 ms immediately before the epoch. All time points after downsampling (625 time points for Epochs 1 and 2 and 500 time points for Epoch 3) and all 32 EEG electrodes were used in the test (i.e, 20000 comparisons for Epochs 1 and 2 and 16000 comparisons for Epoch 3). Repeated measures t-tests were performed for each comparison for the original data, as well as 2500 random within-participant permutations of the original data. For each permutation, t-scores that corresponded to uncorrected p-values of 0.01 or less were combined with neighboring t-scores that also fulfilled the same condition to form clusters. Spatial neighbors were defined as electrodes whose distance was no more than 5.77 cm. Temporal neighbors were adjacent time points. We then calculated the *mass* of all the clusters - the sum of t-scores in that cluster. The most extreme cluster mass (whose absolute value was the largest) for each of the 2501 permutations was recorded, which together formed the null distribution. Using this distribution, p-values were derived for each cluster in the original data, allowing us to derive statistically significant clusters. Two tests were carried out for each epoch using this procedure, one for the difference between sentences with 2 referents and 3 referents and one for the difference between sentences with similar referents and dissimilar referents.

Additionally, we used the Factorial Mass Univariate ERP Toolbox (FMUT) (Fields, 2017) to examine the interaction between the number of referents and the similarity between referents. Fields and Kuperberg (2020) argued that the MUT analysis is only applicable for simple designs and FMUT was designed to make use of mass univariate approaches for factorial ANOVA to investigate complex factorial designs. Our preprocessing procedure prior to the statistical test for the factorial cluster-based permutation test using FMUT was the same as that for the cluster-based permutation test using MUT. The test was also carried out using a family-wise alpha level of 0.01.

4.4.6 Results

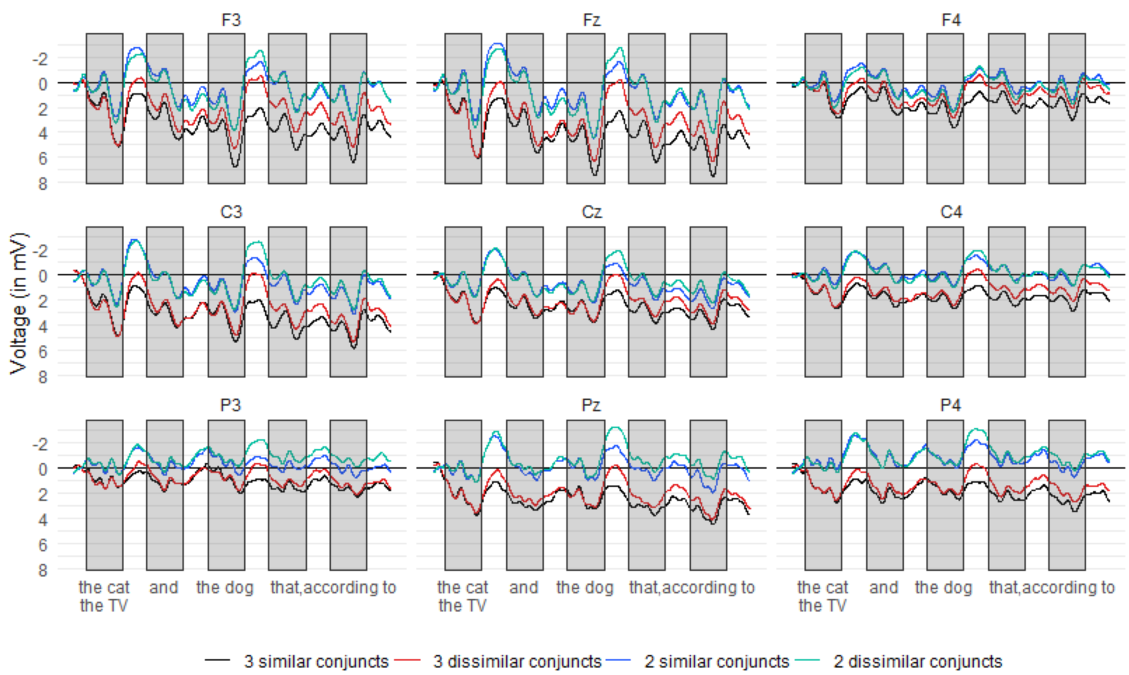

Figure 4.5: Experiment 5: Grand averaged ERPs for Epoch 1 of all conditions

For the 2500-ms Epoch 1, which corresponds to the encoding and early maintenance of the coordinated filler NP into WM, visual inspection of Fig 4.6 showed that there is a sustained negativity where the 2-referent conditions elicited more negativity than the 3-referent conditions for the whole epoch. Like the sustained negativity in Experiment 3, this negativity is global, although both Figs 4.6 and 4.7 demonstrated smaller magnitude for right anterior electrodes and left posterior electrodes. The sustained negativity did not increase or decrease progressively during the course of the epoch although there are significant increases in magnitude in certain time windows, which will be discussed further in the analysis of Epoch 2. The outcome of the cluster-based permutation showed 2 clusters: one cluster than spans the whole epoch and includes most EEG electrodes and one cluster that spans the second half of the epoch and only includes the extreme electrode FT10. The latter cluster is likely an artifact. The former supported the existence of a sustained global negativity between the 2-referent and 3-referent conditions.

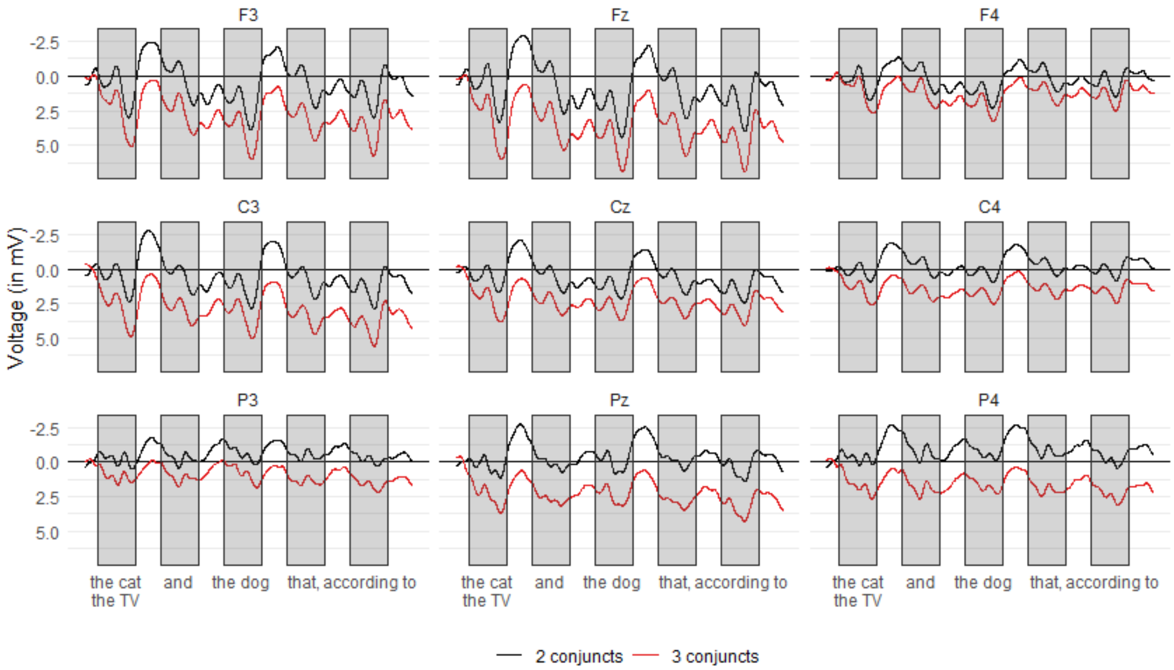

Figure 4.6: Experiment 5: Grand averaged ERPs by number of referents for Epoch 1

The comparison between conditions with similar referents and dissimilar referents in Epoch 1 also revealed a sustained signal. Visual inspection of Figs 4.8 and 4.9 showed that conditions with dissimilar referents elicited more negativity than conditions with similar referents starting around 300 ms after *the dog* and lasting to the end of Epoch 1. The negativity was more pronounced in left anterior electrodes, especially for the 3-referent conditions. Lastly, the amplitude of the negativity was highest at the beginning (300 ms after the onset of *the dog*) and became smaller on the subsequent words *that* and *according to*.

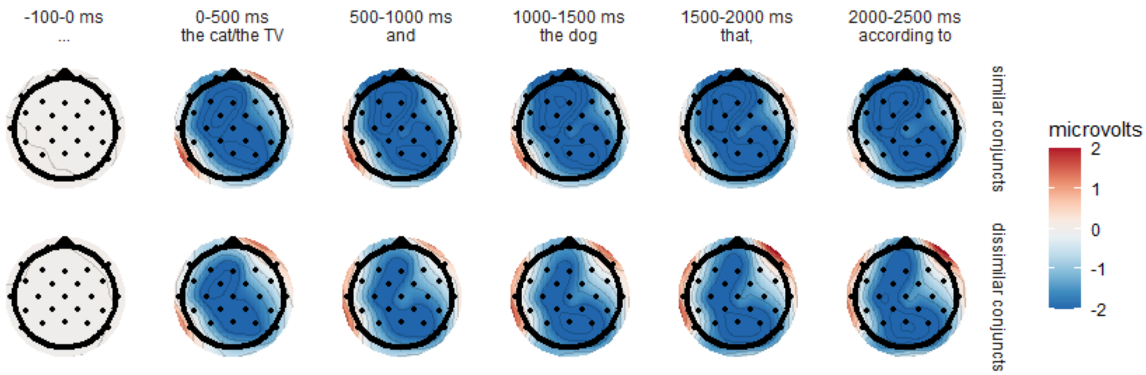

Figure 4.7: Experiment 5: Topograms for 2 referents - 3 referents at both similar and dissimilar conditions for Epoch 1

Clustering analysis corroborated this observation. 3 negative clusters were significant: the first cluster starting 268 ms after the onset of *the dog* and lasting until the onset of *that* and the second and third starting 272 ms after the onset of *that* and lasting until the onset of *according to*. Again, the separation of the second and third clusters was likely artifactual. Thus, these clusters supported the existence of a sustained left anterior negativity starting from *the dog* in dissimilar referent conditions relative to similar referent conditions.

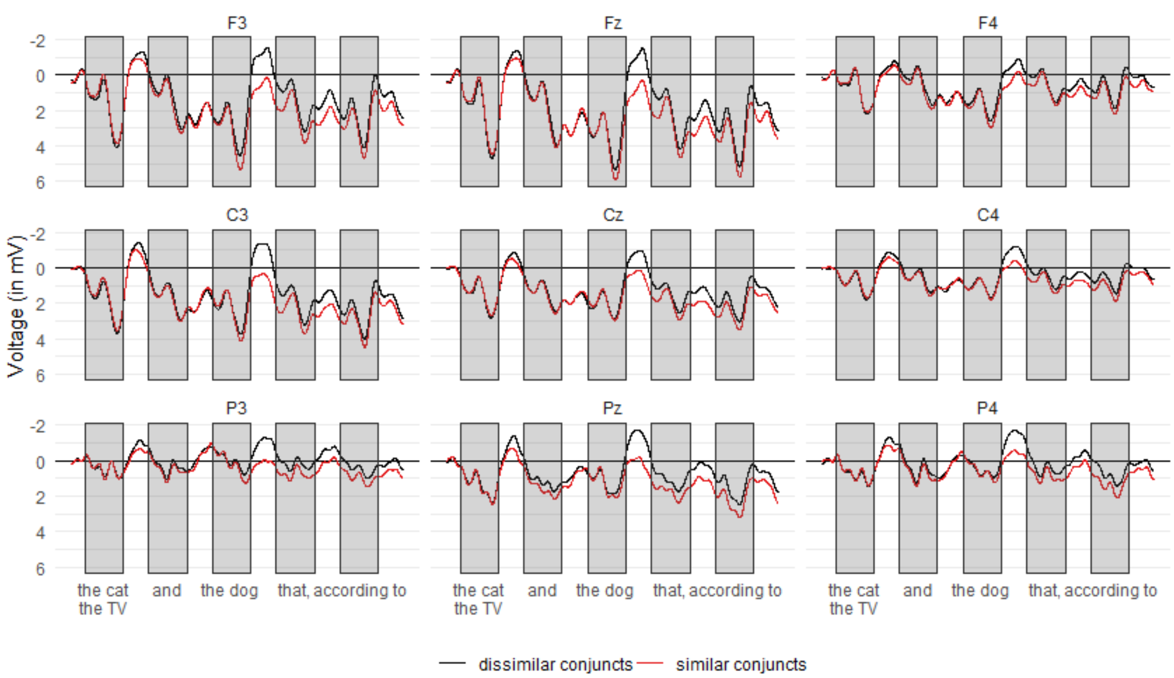

Figure 4.8: Experiment 5: Grand averaged ERPs by referent similarity for Epoch 1

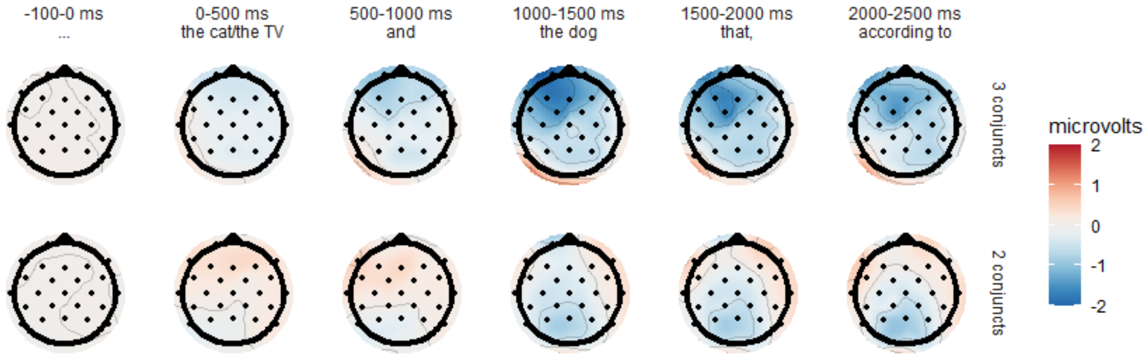

Figure 4.9: Experiment 5: Topograms for dissimilar referents - similar referents at both 2-referent and 3-referent conditions for Epoch 1

Lastly, in Epoch 1, the factorial clustering analysis using FMUT also revealed a significant interaction between number of referents and referent similarity in a cluster starting from 120 ms after *the dog*'s onset and lasting to 396 ms after the onset. Visual inspection of Fig 4.9 showed that while the sustained anterior negativity existed between the 3-referent conditions, it was absent in the 2-referent conditions. This pattern held for *the dog*, *that* and *according to* upon visual inspection but only the time window on *the dog* showed a significant interaction. A post-hoc comparison between the dissimilar and similar 3-referent conditions showed 4 significant negative clusters on four words *the cat/the TV*, *the dog*, *that* and *according to*. The cluster on *the dog* was the largest and covered the most electrode-time point combinations. The clusters on *that* and *the cat/the TV* were smaller and the cluster on *according to* was the smallest. All the four clusters were more concentrated around anterior electrodes. To the contrary, a comparison between the dissimilar and similar 2-referent conditions only showed 1 significant negative cluster on *the dog*. The cluster was much smaller than its counterpart for the 3-referent conditions. Additionally, most electrodes in this cluster were posterior electrodes, as evident by the posterior negativity on *the dog* in Fig 4.9.

The 2500-ms Epoch 2, which spans the maintenance period of the coordinated NP, was analyzed to understand what remained after the large sustained negativity in Epoch 1 was removed after re-baselining.

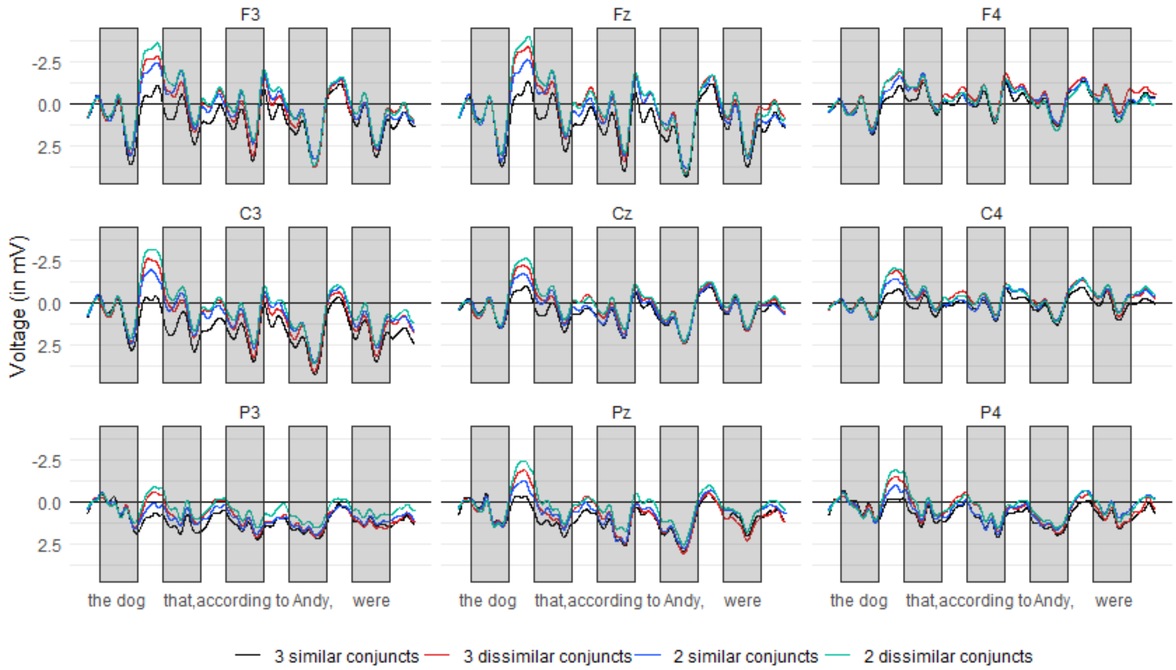

Figure 4.10: Experiment 5: Grand averaged ERPs for Epoch 2 of all conditions

For the referent similarity comparison, visual inspection of Figs 4.13 and 4.14 showed similar patterns as the analysis in Epoch 1. The dissimilar-referent conditions, especially the one with 3 referents, showed negativities relative to the similar-referent conditions on *the dog*, *that* and *according to*. These negativities were concentrated around left anterior electrodes and the magnitude decreased progressively. Accordingly, the clustering analysis found 3 significant negative clusters (refer to table 4.3): one beginning 244 ms after the onset of *the dog* and persisting for 384 ms, and two clusters beginning 300 ms after *that* and persisting for 336 ms. The latter two clusters were likely part of a larger cluster. The electrodes that were significant in the most time points for these clusters were left anterior electrodes and midline electrodes (with the exception of Oz). These results reinforced those observed in the analysis of Epoch 1.

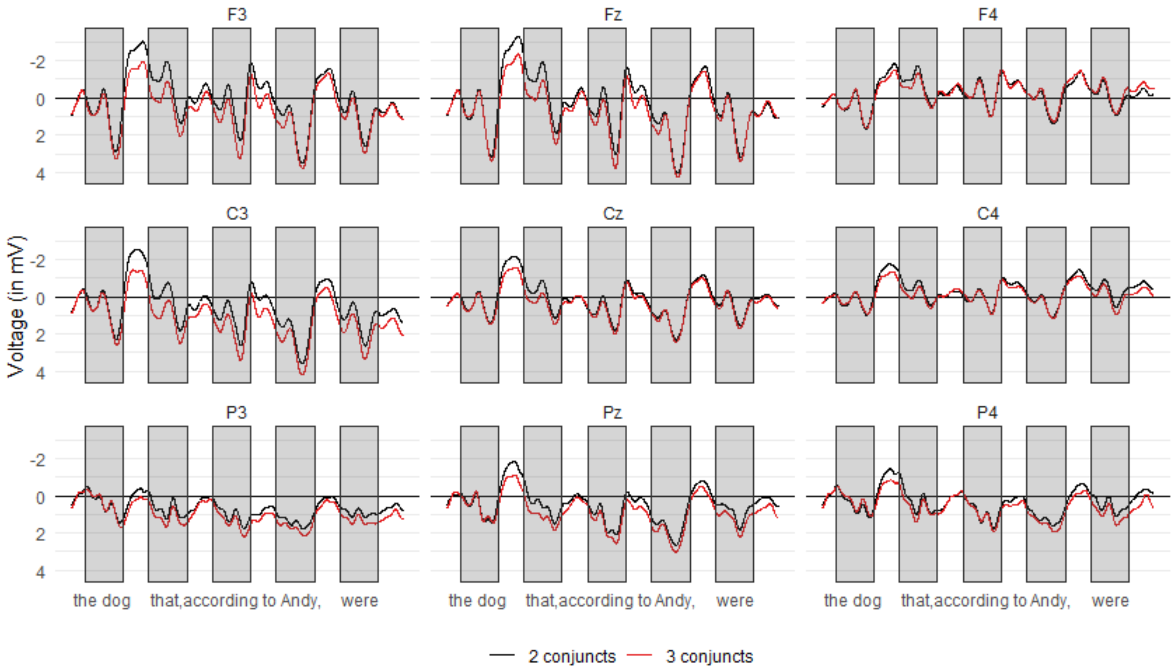

Figure 4.11: Experiment 5: Grand averaged ERPs by number of referents for Epoch 2

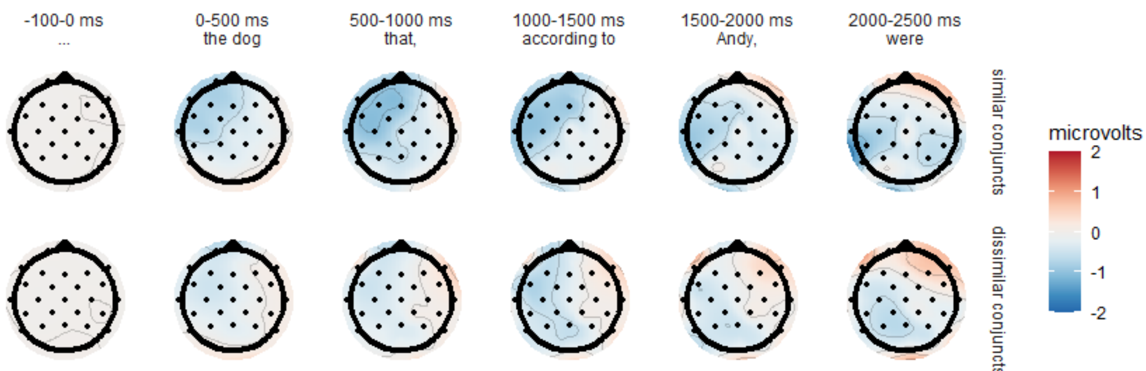

Figure 4.12: Experiment 5: Topograms for 2 referents - 3 referents at both similar and dissimilar conditions for Epoch 2

The comparison between 2-referent and 3-referent conditions in Epoch 2 shared similarities with the comparison between dissimilar-referent and similar-referent conditions. Visual inspection of Figs 4.11 and 4.12 revealed showed negativities on *the dog*, *that* and *according to* in the 2-referent conditions relative to the 3-referent conditions. The magnitude of the negativity also decreased progressively across the time period. The clustering analysis confirmed that the 2-referent conditions elicited more negativity than the 3-referent conditions in this region. Two significant negative clusters were detected: one starting 264 ms after the onset of *the dog* and lasting for 428 ms and one starting 112 ms after the onset of *according to* and lasting for 308 ms. No significant cluster was detected on the word *that*. Additionally, the clustering analysis showed no significant interaction between referent similarity and number of referents in this epoch.

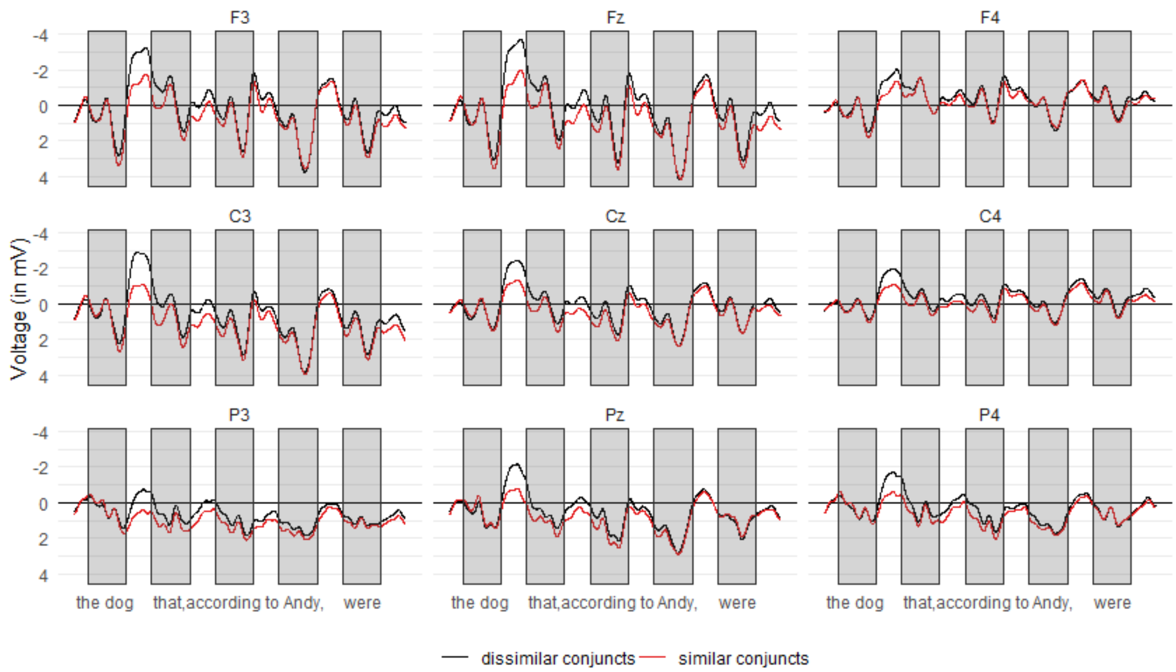

Figure 4.13: Experiment 5: Grand averaged ERPs by referent similarity for Epoch 2

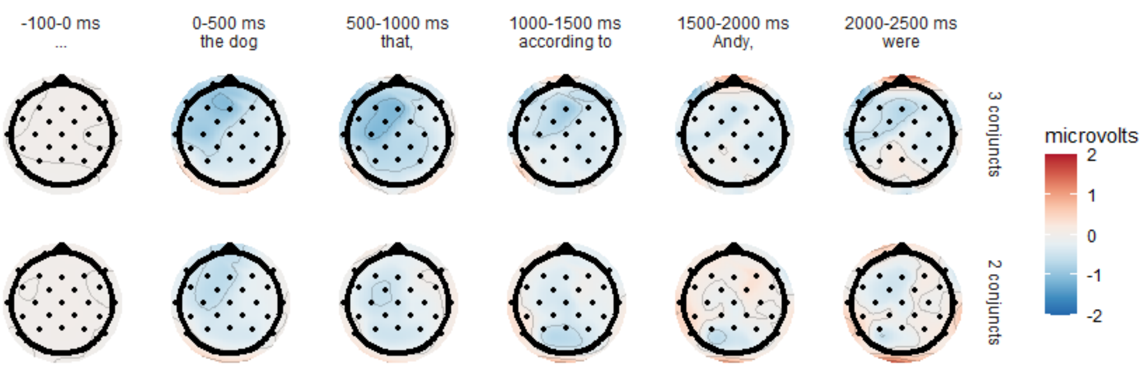

Figure 4.14: Experiment 5: Topograms for dissimilar referents - similar referents at both 2-referent and 3-referent conditions for Epoch 2

For the 2000-ms Epoch 3, which corresponds to the retrieval of the coordinated filler NP from WM, the clustering analysis revealed no significant clusters for both comparisons, which corroborated the lack of visible patterns in Figs 4.16 and 4.18.

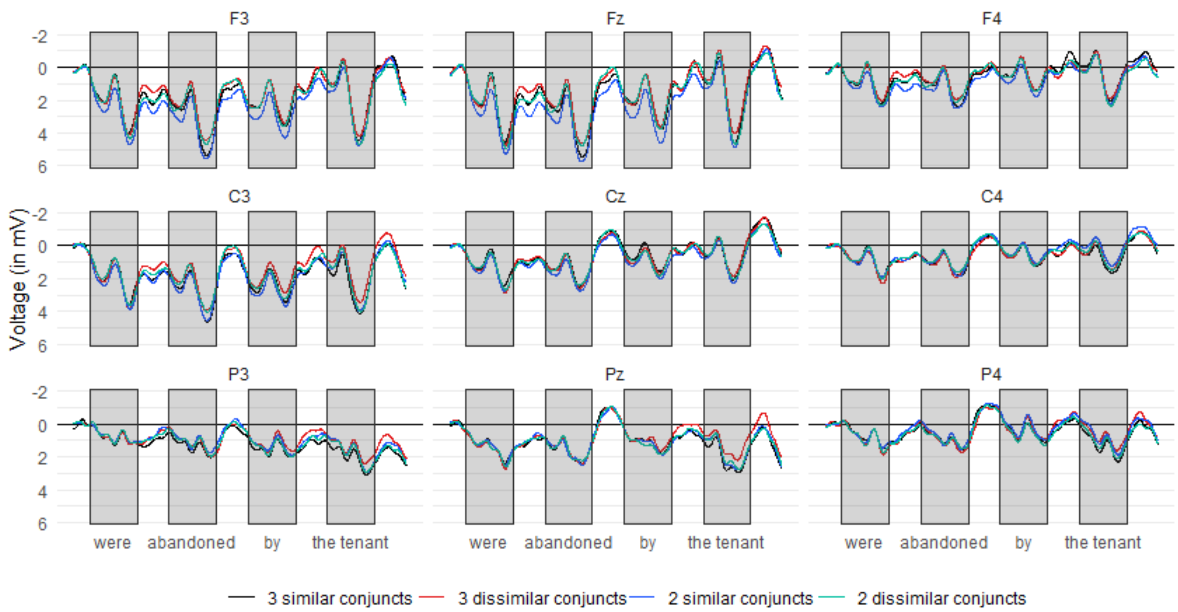

Figure 4.15: Experiment 5: Grand averaged ERPs for Epoch 3 of all conditions

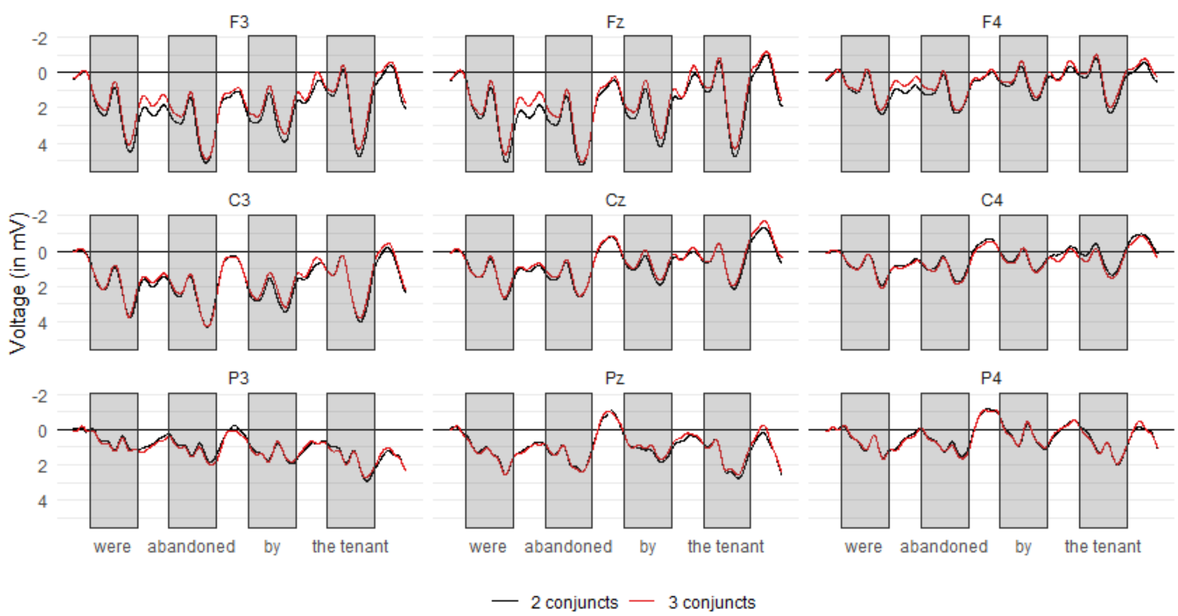

Figure 4.16: Experiment 5: Grand averaged ERPs by number of referents for Epoch 3

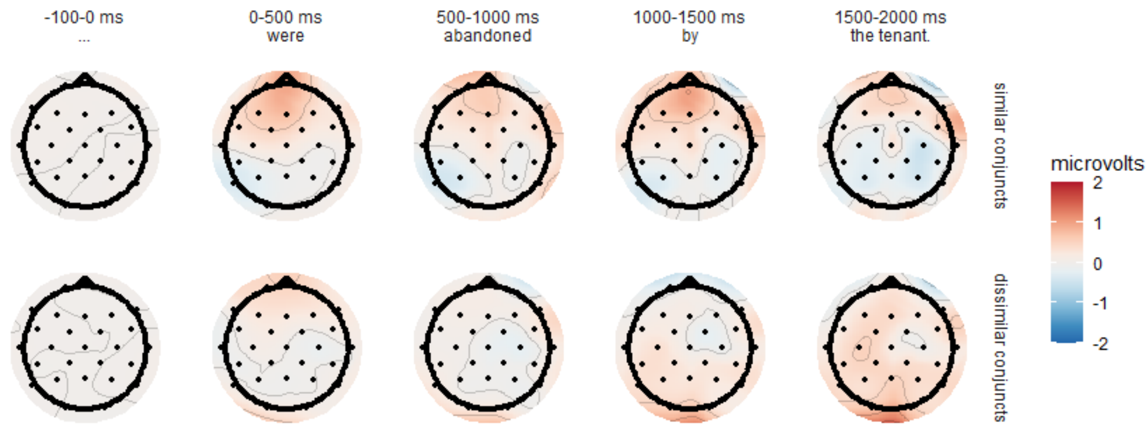

Figure 4.17: Experiment 5: Topograms for 2 referents - 3 referents at both similar and dissimilar conditions for Epoch 3

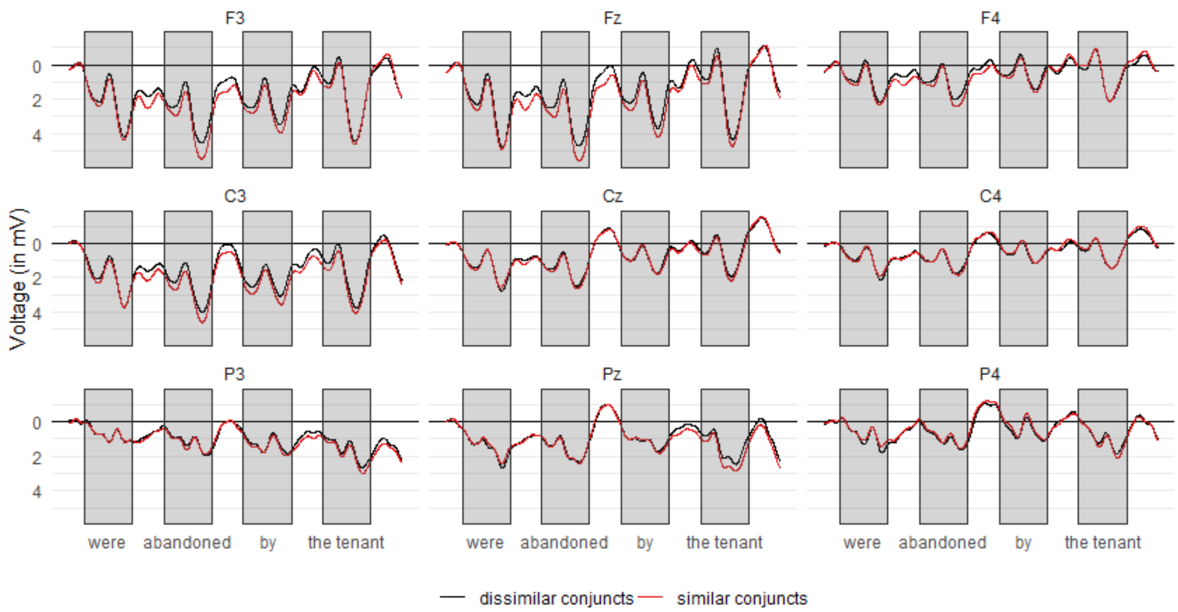

Figure 4.18: Experiment 5: Grand averaged ERPs by referent similarity for Epoch 3

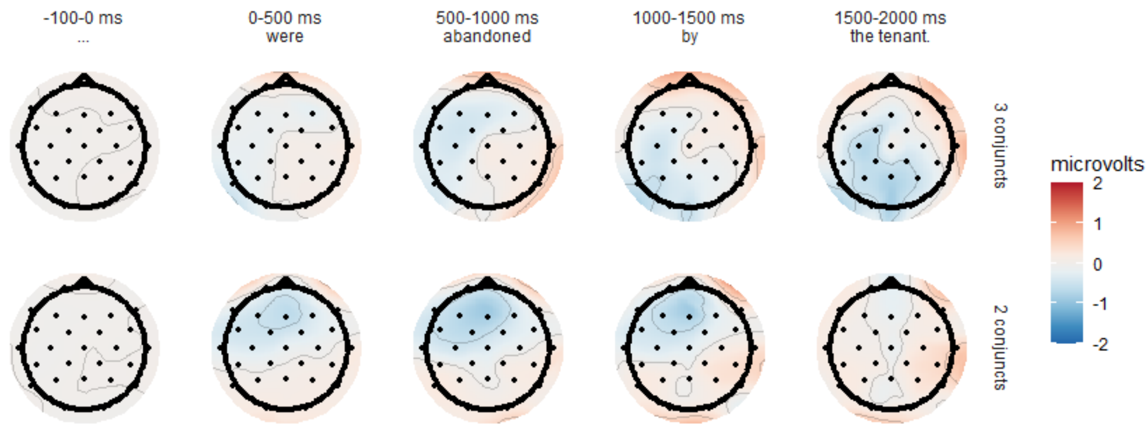

Figure 4.19: Experiment 5: Topograms for dissimilar referents - similar referents at both 2-referent and 3-referent conditions for Epoch 3

4.4.7 Discussion

In this experiment, our goal was to investigate the effect of filler NP's representational complexity and internal coherence, as well as their interaction on its WM encoding, maintenance and retrieval through an ERP lens. As noted in section 4.3.4, some drawbacks with behavioral RT measures include: 1) lack of understanding for the maintenance period given few behavioral results during the maintenance site and 2) RT spillover made it difficult to determine whether an effect during the maintenance site was due to difference in maintenance costs or spillover due to difference in encoding costs. In contrast, in the EEG literature, the SAN is a relatively well-characterized ERP index that appears to reflect difference in maintenance costs (see Section 3.1.2 for more discussion about the evidence for and against this functional interpretation of the SAN). We can thus leverage the knowledge of the SAN to investigate the effects of representational complexity and internal coherence of the filler NP at the maintenance site. Additionally, the synthesis of both behavioral results from Experiments 4a and 4b and neural results from Experiment 5 can give us a multidimensional view of the WM processes involving complex linguistic representations.

For the filler NP's representational complexity contrast, we manipulated the number of referents in the coordinated filler NP. Unlike in Experiments 4a and 4b which had a three-way contrast (1, 2 and 3 referents), Experiment 5 only compared between 2-referent

Epoch	Effect	Timing	Electrode
1	Number of referents (2 clusters)	0-2496 ms	FP1, Fz, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, Pz, P3, P7, O1, Oz, O2, P4, P8, TP10, CP6, CP2, Cz, C4, T8, FT10, FC6, FC2, F4, F8, FP2
		1192-2496 ms	FT10
	referent similarity (3 clusters)	1268-1584 ms	FP1, Fz, F3, F7, FT9, FC5, FC1, C3, T7, CP5, CP1, Pz, P3, P7, P4, CP6, CP2, Cz, C4, FC2, F4, FP2
		1772-1860 ms	Fz, F3, FC1, C3, CP1, Pz, P3, P4, CP2, Cz, C4, FC2
		1860-2044 ms	Fz, F3, FC1, C3, CP1, Pz, P3, O2, P4, CP2, Cz, C4, FC2, F4
	Number x Similarity (1 cluster)	1120-1396 ms	FP1, Fz, F3, F7, FC5, FC1, T7, TP9, P7, O1, Oz, O2, C4, FC6, FC2, F4, F8, FP2
2	Number of referents (2 clusters)	264-692 ms	FP1, Fz, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, Pz, P3, P4, CP2, Cz, C4, FC2, F4, FP2
		1112-1420 ms	FP1, F3, F7, FC5, FC1, C3, T7, TP9, CP5, CP1, P7, O1
	referent similarity (3 clusters)	244-628 ms	FP1, Fz, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, Pz, P3, P4, P8, TP10, CP6, CP2, Cz, C4, FT10, FC6, FC2, F4, F8, FP2
		800-876 ms	Fz, F3, FC5, FC1, C3, T7, CP1, Pz, P3, P4, CP2, Cz, C4, FC2
		908-1036 ms	Fz, F3, FC5, FC1, C3, CP1, Pz, P3, P4, CP6, CP2, Cz, C4, FC2, F4

Table 4.3: Experiment 5: Timing and electrodes of significant clusters from FMUT analysis

and 3-referent conditions. Epoch 1, where the filler NP was being encoded into WM and maintained there, showed a sustained global negativity in the 2-referent conditions relative to the 3-referent conditions that spanned the whole epoch. This sustained negativity was akin to that in Epoch 1 of Experiment 3. They were both present in the conditions with less complex filler NPs (NPs that were not pre-modified in Experiment 3 and NPs that had fewer conjoined referents in Experiment 5), relative to more complex counterparts (pre-modified NPs in Experiment 3 and NPs that had more conjoined referents in Experiment 5).

Hence, the interpretation of this negativity is similar to that discussed in Chapter 3. It could represent the increased maintenance cost associated with maintaining a less complex representation. The major difference between Epoch 1 in Experiment 3 and Epoch 1 in Experiment 5 is that in this experiment, Epoch 1 spanned several words in the encoding site (*the cat/TV and the dog*) while in Experiment 3, Epoch 1 only included the last word of the encoding site. The existence of the negativity early during encoding could be due to the concurrence of WM encoding and maintenance. Since language comprehension is incremental, while new referents were being read and encoded into WM as part of the NP, part of the NP was being maintained in WM. Thus, the sustained negativity during the encoding site could index difference in WM maintenance cost.

However, like Experiment 3, it is possible that this global negativity was generated due to different pre-baseline materials between the 2- and 3-referent conditions (Steinhauer and Drury, 2012). In the 2-referent conditions, the preceding region contained the cleft *It was* while in the 3-referent conditions, it contained the first referent *the hamster/the cat*. Since these pre-baseline materials were very different lexically and grammatically, they could have generated a transient ERP divergence that would be over-corrected by the re-baselining algorithm. In fact, both complex conditions in Experiment 3 and 5 were preceded by content words (nouns or adjectives) while both simple conditions were preceded by more functional words (the cleft *it was* or the complementizer *that*). Since content words are known to generate greater N400 effect than function words (Garnsey, 1986; Kutas and Hillyard, 1983), there was an existing transient negativity pre-baseline in the complex conditions that was corrected by pre-baselining, possibly leading to sustained negativity in the simple conditions, as observed in Epoch 1 of Experiment 3 and Experiment 5. This design shortcoming made the referent number comparison in Epoch 1 hard to interpret.

In Epoch 2, which included the end of encoding, maintenance and the start of retrieval, we found transient negativities in the 2-referent conditions (relative to the 3-referent conditions) that started after the onset of *the dog* and *according to*. Visual inspection of Fig 4.12 suggested that there was also a transient negativity after the onset of *that* but only for the similar-referent conditions. Critically, these transient negativities were elicited in left anterior electrodes, which are the same topographical distribution of the SAN. Visual inspection of Fig 4.11 revealed that on these three words, the negativities arose during the blank space after the presentation of the words and decreased in magnitude when the next word was presented. This is similar to some previous manifestations of the SAN, such as in Lau and Liao (2018) and in Experiment 1 of Cruz Heredia et al. (2022), as shown in Figs 4.20 and 4.21, respectively. In these instances of the SAN, they also observed sustained negativities that decreased in magnitude during word presentation. Since these experiments relied on ANOVA on averages of whole time windows for statistical analysis, these transient negativities that decreased in magnitude and increased again after word presentation could appear as a sustained negativity. In contrast, our analysis, which made use of spatiotemporal clustering, had the advantage of pinpointing more detailed onsets and offsets of broadly distributed effects (Groppe et al., 2011). However, the goal of this comparison is to show that although the ERP signal we observed appears as consecutive transient negativities, it resembled some manifestations of the SAN in the literature in topographic distribution and behavior with regards to changes in magnitude.

With this consideration, our data from Epoch 2 showed a SAN in 2-referent conditions relative to 3-referent conditions which lasted from the onset of *the dog* to the onset of *Andy*. This suggested that having a more complex NP which comprises more conjoined referents facilitates WM maintenance, an implication which aligned with the results of behavioral experiments 4a and 4b. Critically, the extent of this facilitation effect mirrored that in the behavioral experiments. In all behavioral experiments, 3-referent conditions were read faster than 2-referent conditions on the region *according to* and not on the subsequent region *Mary*. In this experiment, the SAN lasted until the onset of *Andy* where it dropped off. Furthermore, we did not observe a difference between the 2- and 3-referent conditions on the word *that* in Experiment 4b, which also juxtaposed the weaker SAN on that word (which was only observed in visual inspection and not in the clustering analysis). It is unclear why the effect weakened on the word *that* in both experiments. The parallels between the behavioral and neural experiments also extended to the retrieval site. In the behavioral experiments, we observed no significant effect of referent number in this area (despite there being weak evidence for a few effects) and in this ERP experiment, we observed no significant clusters in Epoch 3, which spanned the retrieval site.

For the filler NP's referent similarity contrast, the pattern was similar to that observed in Epoch 2 for the number of referents contrast. Analyses of both Epochs 1 and 2 showed two similar negative clusters for the dissimilar-referent conditions relative to the similar-referent conditions: one after the onset of *the dog* and one after the onset of *that*. Visual inspection of both Figs 4.9 and 4.14, as well as the post-hoc analysis of 3-referent and 2-referent conditions separately, revealed two additional negative clusters on *the cat/the TV* and *according to* only for the 3-referent conditions, not for the 2-referent conditions. For the transient negativities appearing on *the dog*, *that* and *according to*, visual inspection of Fig 4.8 suggests that they made up a SAN, just like the transient negativities detected in the referent number contrast in Epoch 2. They were also anterior and slightly left in distribution, akin to previous reports of the SAN. This appearance of the SAN in this region demonstrates that the maintenance of less coherent NP in WM is more costly than that of more coherent NP in WM.

Again, similar to the number contrast, the results of this experiment also resembled those of the behavioral experiments. The behavioral experiment 4b showed a facilitation effect of referent similarity on *the dog* and *that*, which also were present as negative clusters in Experiment 5. Additionally, none of the three experiments observed a significant effect during the retrieval site (Epoch 3 of this experiment). Altogether, number of referents and referent similarity both consistently influenced the encoding and maintenance processes and to a much lesser extent, the retrieval process.

It is unclear if the negativity detected on *the cat/the TV* between the 3 dissimilar-referent and 3 similar-referent conditions was part of the SAN, especially since no significant clusters were found on the word *and* which linked the referents. It is more likely an N400 effect which resulted from the second referent being a less expected continuation in the dissimilar conditions (Kutas and Hillyard, 1984). This might also explain why the negativity after the onset of *the dog* was larger in magnitude, distribution and duration than other negativities detected. This was likely the end result of the combination between the onset of the SAN and an even bigger N400 effect than that on *the cat/the TV* since in the 3 dissimilar-referent condition, this referent was different in categories from the other two referents. This bigger N400 effect was also evident in the significant interaction between number of referents and referent similarity after the onset of *the dog*. The interaction resulted from 3-referent conditions having a larger contrast due to referent similarity than the 2-referent conditions. The reason behind this discrepancy is possibly the larger N400 effect in the 3 dissimilar-referent condition than in the 2 dissimilar-referent condition. The co-occurence of the N400 and the SAN showed that internal coherence of the filler NP both facilitates word access and WM maintenance.

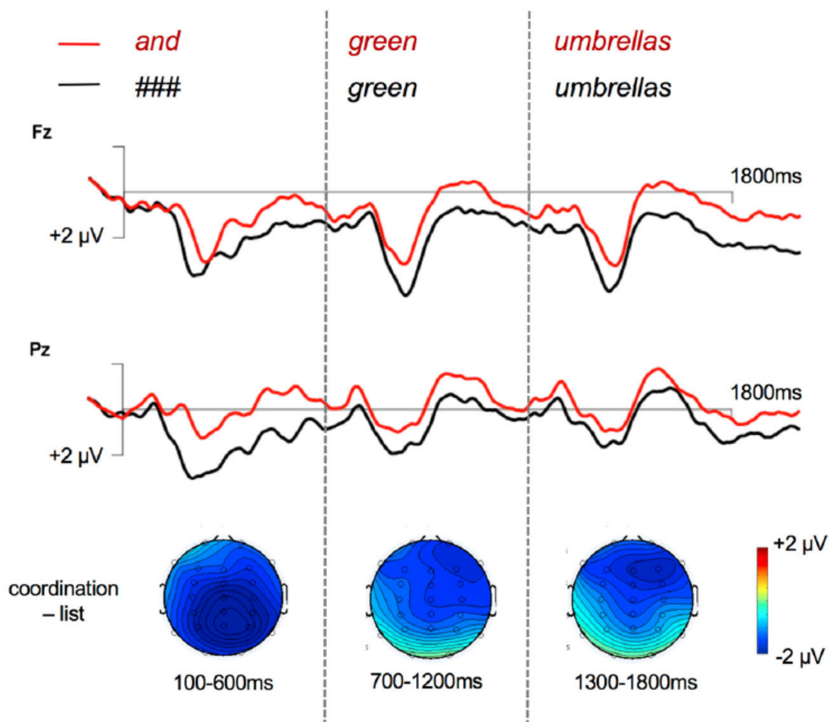

Figure 4.20: SAN observed by comparing coordinated NPs and list of NPs. SOURCE: Lau and Liao (2018).

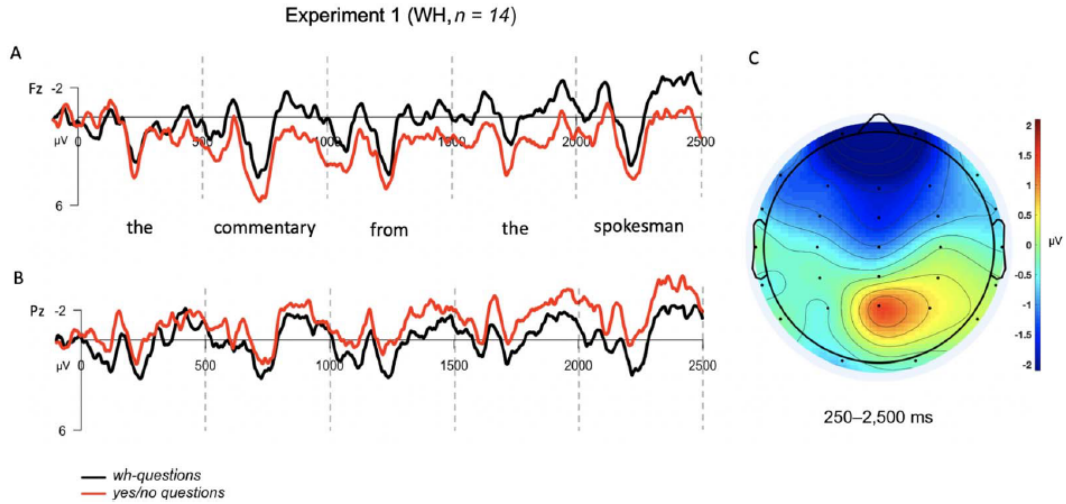

Figure 4.21: SAN observed by comparing matrix *wh-* questions and *yes-no* questions. SOURCE: Cruz Heredia et al. (2022). Reproduced with permission from *MIT Press*.

4.5 General Discussion

Through Experiments 4a, 4b and 5, our goal was to investigate how the internal coherence of the filler NP modulates its encoding, maintenance and retrieval through both a behavioral and neural lens. Furthermore, we examined how these effects interact with different levels of representational complexity of the NP. Lastly, Experiment 5, which observed two different instances of the SAN, can allow us to contribute to the discussion on the functional interpretation of the SAN.

4.5.1 On Internal Coherence

To study internal coherence, we manipulated how similar the referents within a coordinated NP are by selecting referents that either all belong to the same conceptual category or all belong to different categories. During encoding, reading more coherent NPs, relative to reading less coherent NPs, reduced processing cost, particularly on the final word of the NP *the dog*, as demonstrated by both faster RTs in behavioral experiments and less negativity in the EEG experiment. During early maintenance, before encountering another nominal referent *Mary*, we observed a similar facilitatory effect, again shown by faster RTs and lower SAN in conditions with more coherent NPs. During late maintenance from *Mary* and retrieval, there was weak evidence for some facilitatory effects on RT, showing that retrieval facilitation thanks to high internal coherence is not as robust as those observed during encoding and maintenance. Overall, the experiments showed that increased internal coherence of the target NP is always beneficial to the WM processes, especially encoding and maintenance.

These results are consistent with the findings of Hofmeister (2011)'s Experiment 3, where the encoding and retrieval of a more coherent NP were faster than its incoherent counterpart. Our experiments showed that this effect is generalizable between different kinds of coherence, whether it be feature coherence between the noun and its modifiers like in Hofmeister (2011) or between the referents in a coordinated NP like in our experiments. Our results also echoed Hofmeister (2011) that more encoding time does not always lead to reduced WM cost for later WM processes of maintenance and retrieval. However, the major discrepancy between our results and Hofmeister (2011)'s is the lack of robust retrieval site's effects in our experiments, both behaviorally and neurally. One possible reason is that facilitation during the maintenance site pre-empts the difficulty of retrieving inconsistent target NP. Thus, the cost of retrieving incoherent features is front-loaded by organizing the features more efficiently during encoding and early maintenance. This explanation is unlikely given that Hofmeister (2011)'s materials also had an extended maintenance period (*who the activist looking for more contributions*) but still detected retrieval site's facilitation. Another alternative but related explanation is that the intervening material in our maintenance period, a parenthetical, is less demanding to memory than that of Hofmeister (2011), since the retrieval site was in a restrictive relative clause. The increased demand of a relative clause might lead to increased importance from increased coherence in the target NP. Furthermore, referents introduced in parentheticals have been shown to participate less in interference during retrieval (Dillon et al., 2014). Thus, since our stimuli introduced an intervening referent *Mary* in a parenthetical, there might have been less interference at the retrieval site, leading to less benefit from having a more coherent target NP.

Our results also confirmed our hypothesis that contrary to instances of similarity-based interference (Lewis and Vasishth, 2005), feature similarity between referents that will be retrieved together actually benefits WM processes. This pattern was observed in the visual WM literature, where similarity is inhibitory if the items are retrieved individually but is facilitatory if the items are retrieved as an ensemble (Mate and Baqués, 2009). It is possible that akin to the visual domain, having referents that share features can make grouping them into a representation more efficiently. Higher coherence also makes storing the representation less costly since fewer features need to be stored if the component referents share more features. Other possible accounts for the facilitatory effect of similarity, such as sharper representations or less drifting to prototypical value (Lin and Luck, 2009), are less appealing to explain our results since unlike colors of visual objects, linguistic features are more categorical.

Lastly, our results also supported the theory that semantic features of referents are not “actively maintained" in WM (Ness and Meltzer-Asscher, 2017,1; Wagers and Phillips, 2014). Our experiments showed that while there was a facilitatory effect of coherence during early maintenance, the effect was not present during late maintenance (on the noun *Mary*). Since internal coherence is a measure of the relationship between semantic features, this trend suggested that the memory of semantic features “decayed" and was not present on the word *Mary*. The SAN observed between dissimilar and similar referent conditions in Experiment 5 was the largest on *the dog* and slowly decreased in magnitude, supporting further the view that semantic features are not actively maintained. However, our results also did not rule out the possibility that the semantic features do not decay but are transferred to activated LTM, particularly since a new referent *Mary* is being introduced. This interpretation is also consistent with the lack of a sustained signal during late maintenance, since the presence of a sustained signal was thought to index WM load, not materials in activated LTM (Foster et al., 2019). Further work is needed to disentangle between these possibilities.

4.5.2 On Representational Complexity

Though Experiments 4a, 4b and 5 were designed to examine the effect of internal coherence on WM processes, the use of coordinated NPs also allowed us to further explore how representational complexity of the target NP modulates the WM processes and how these effects interact with the effects of the target NP's internal coherence. Representational complexity was manipulated by changing the number of conjoined referents in the target NP. Experiments 4a and 4b compared between conditions where the target NP had no conjunction, two conjuncts and three conjuncts while Experiment 5 only contained conditions where the target NP had two conjuncts or three conjuncts.

A common observation between these experiments is that two-referent NPs required less encoding effort but more maintenance effort (and to a lesser extent, retrieval effort) than both one-referent and three-referent NPs. Since three-referent NPs are more representationally complex than two-referent NPs, this pattern agreed with the results in Experiments 1a and 1b, where the encoding of more complex NPs required more effort but their maintenance and retrieval required less effort than their simpler counterparts. The higher encoding effort associated with one-referent NPs, however, with the results of Experiments 2a and 2b. Unlike Experiments 2a and 2b, this effect cannot be explained by the presence of priming in the encoding of two-referent NPs since the first referent might be similar or dissimilar from the second referent. While a referent might prime a similar second referent, the priming would not extend to dissimilar second referents.

Though we still do not understand why one-referent NPs did not behave as expected from simple NPs, a crucial pattern that we observed from these experiments with regards to representational complexity is that maintenance and retrieval effort was inversely proportional to encoding effort. This pattern provided strong support for the time-dependent attention account proposed by Karimi et al. (2020), who hypothesized that more time spent encoding helps recruit more attention and/or memory resources to more complex NPs. The increased resources dedicated to them can increase their activation level and ease their maintenance and retrieval.

4.6 Conclusion

In a series of two SPR experiments and an EEG experiment, we investigated how the internal coherence and representational complexity of target filler NPs affect the relative effort associated with their encoding, maintenance and retrieval. Our experiments revealed that having representations whose features are coherent can facilitate both encoding and early maintenance, possibly by enabling more efficient organization of features. However, this benefit becomes progressively weaker into the later maintenance and retrieval sites, suggesting that semantic features of the target representations have decayed or have been transferred to LTM. Additionally, the comparison between sentences with complex and simple target NPs showed that the facilitatory effect observed during maintenance are dependent on the amount of effort spent encoding the target NP.

CHAPTER 5
GENERAL DISCUSSION AND CONCLUSION

The final chapter will bring back and synthesize the insights glimpsed from the eight experiments detailed in the previous chapters. I will start by summarizing what we found from the experiments and then discuss how these experiments contribute to what we know about WM processes in sentence comprehension and to discussions about the interpretations of sustained neural signals in sentence comprehension. Finally, I will discuss some limitations from the current set of experiments and how they can be improved upon in future studies.

5.1 Summary of Experiments

The dissertation consisted of 8 experiments, 6 behavioral SPR experiments and 2 neural EEG experiments. Experiments 1a, 1b, 2a, 2b and 3 examined how representational complexity of the target NPs affects the WM processes. Specifically, Experiments 1a and 1b studied NPs pre-modified by adjectives or nouns as instances of complex NPs while Experiments 2a and 2b studied coordinated NPs as complex NPs. All these four experiments were done using the SPR paradigm. Experiment 3 was an EEG experiment comparing the ERP signals elicited when participants read sentences with complex pre-modified NPs and simple non-pre-modified NPs. Experiments 4a, 4b and 5 examined how internal coherence and representational complexity of the target NPs influence the WM processes. Experiments 4a and 4b were SPR experiments while Experiment 5 was an EEG experiment. All three experiments investigated processing of sentences with target NPs that varied in the number of coordinated referents and the similarity between those referents.

Experiments 1a and 1b were two SPR experiments that asked how the encoding, maintenance and retrieval of complex pre-modified target NPs and simple non-pre-modified target NPs differ and what processes motivate these differences. Both experiments had a 2 x 2 design where we manipulated the complexity of the target NP and the type of RC the sentence contained. The target NP complexity manipulation involved varying whether the target filler NP was pre-modified by two adjectival or nominal modifiers (complex, *those* *emotional crash survivors*) or not (simple, *those survivors*). In addition, the site where the target NP was retrieved, the RC verb, could be situated in either an SRC or an ORC, which varied the number of competing referents with the target NP. In Experiment 1a, there was no competing referent to the target NP at the retrieval site if the sentence had an SRC. To the contrary, if the sentence had an ORC, there was one competing referent at the retrieval site - the RC subject NP. In Experiment 1b, we added an additional layer of embedding and thus, an additional competing referent to all conditions. As a result, there was one or two competing referents if the sentence had an SRC or ORC, respectively. In both experiments, we examined the RTs at the target NP's head noun (encoding site), the RC verb and its spillover regions (retrieval site) and the words between them (maintenance site).

Experiments 2a and 2b were two SPR experiments that asked how the encoding, maintenance and retrieval of complex coordinated target NPs and simple uncoordinated target NPs differ and whether these distinctions replicated those observed in Experiments 1a and 1b. Experiments 2a and 2b had the same stimuli and design as Experiments 1a and 1b, respectively, except for the target NP complexity manipulation. In Experiments 2a and 2b, the target NP complexity manipulation involved varying whether the target filler NP was coordinated with two referents (complex, *those judges and lawyers*) or uncoordinated (simple, *those lawyers*).

Experiment 3 was an EEG experiment that examined the encoding, maintenance and retrieval of complex pre-modified target NPs and simple non-pre-modified target NPs through a neural lens. It had similar stimuli and the same 2 x 2 design as Experiment 1a, where we manipulated whether the target NP was pre-modified and whether the RC verb was in an SRC or ORC. However, instead of self-paced reading, sentences were presented using an RSVP paradigm while participants had their EEGs measured. We examined three epochs in Experiment 3: Epoch 1 and 2 which spanned the encoding and maintenance of the target NP and Epoch 3 which spanned its retrieval and several subsequent words.

Experiments 4a and 4b were SPR exepriements that investigated how the internal coherence and representational complexity of the target NP affected its encoding, maintenance and retrieval. Both experiments had a 3 x 2 design where we manipulated the number of referents in the target NP and whether the referents were similar to each other. For the number of referents manipulation, we varied whether the target filler NP was uncoordinated, was coordinated with two referents or with three referents. The degree of representational complexity increased with the number of referents. For the referent similarity manipulation, we varied whether the referents in coordinated NPs were from the same conceptual category or different conceptual categories. In both experiments, we examined the RT at the target NP (encoding site), the matrix VP (retrieval site) and the words between them (maintenance site)

Experiment 5 was an EEG experiment that followed up on Experiments 4a and 4b through a neural lens. Experiment 5 had similar stimuli to Experiment 4b but had a 2 x 2 design. In Experiment 5, the target NP was always coordinated and had either two or three conjoined referent. The referent similarity manipulation remained the same. Instead of self-paced reading, sentences were presented using an RSVP paradigm while participants had their EEGs measured. We examined three epochs in Experiment 5: Epoch 1 which spanned most of the encoding and early maintenance of the target NP, Epoch 2 which spanned late encoding and the entirety of maintenance and Epoch 3 which spanned the retrieval period.

Below is a synthesis of the main findings from the eight experiments in terms of how the experimental manipulations affected each of the three WM processes:

Encoding. The experiments showed that more complex target representations sometimes required more effort to be encoded. Similar to the results observed by Hofmeister (2011), encoding of pre-modified NPs in Experiment 1b took a longer time than that of their simpler non-pre-modified counterparts. Similarly, encoding of three-referent NPs in Experiment 4a and 4b took a longer time than that of their simpler two-referent counterparts. However, this effect was not consistently observed across the experiments. Experiment 1a failed to produce any effect of representational complexity in the encoding site while Experiments 2a, 2b and 4b showed that more complex two-referent coordinated NPs were encoded faster than their simpler uncoordinated counterparts. It is unclear what is responsible for the reverse pattern when comparing between uncoordinated and two-referent NPs. Nevertheless, the other cases supported increased encoding effort for more complex NPs but this effect was not robust.

More internally coherent representations required less effort to be encoded than their less coherent counterparts. This was shown consistently by the decreased RTs in the encoding site of Experiments 4a and 4b when the target NPs consisted of similar referents.

Maintenance. The experiments showed that more complex target representations consistently required less effort to be maintained. Experiments 1a, 1b, 2a, 2b, 4a and 4b all demonstrated shorter RTs in certain words of the maintenance site when the target NPs were more complex (when they were pre-modified or had more conjoined referents). To reinforce this trend, in Experiments 3 and 5, trials with more complex target NPs elicited less global negativity than trials with less complex target NPs during the maintenance period. Additionally, 2-referent trials in Experiment 5 also elicited the SAN relative to 3-referent trials in Epoch 2. Weaker negativities in trials with more complex NPs, along with decreased RTs during maintenance, supported the facilitatory effect of maintaining more complex NPs.

Furthermore, the experiments' results also hinted that there are at least two sources accounting for the facilitatory effect during maintenance of complex NPs: increased distinctiveness between competing referents and increased encoding time. Firstly, complex target NPs have more features, which help distinguish themselves from competing referents and decrease the extent of similarity-based interference. The evidence for this increase in distinctiveness came from stronger facilitation effect from having complex NPs when there were competing referents. Facilitation was stronger in sentences with ORCs (which had a competing referent) in both Experiments 1a and 2a than sentences with SRCs (which had no competing referents). This stronger facilitation disappeared in Experiments 1b and 2b, where both sentences with ORCs and SRCs contained competing referents during maintenance.

Secondly, complex target NPs are longer and require more sheer encoding time, which can make them more salient (Hofmeister, 2011) or recruit more attentional resources towards them (Karimi et al., 2020), resulting in higher activation level and ease of maintenance. The evidence for this came from the presence of facilitatory effect during maintenance even when there were no competing referents, such as in SRCs in Experiment 1a or before the appearance of competing referents, such as in Experiments 4a and 4b. Additionally, in Experiments

4a and 4b, the effects during maintenance mirrored the effects during encoding. Trials with 2-referent target NPs required less encoding effort but more maintenance effort than both trials with uncoordinated or 3-referent target NPs, supporting the idea that higher encoding effort can facilitate the maintenance process. Furthermore, the lower sustained negativities associated with more complex NPs in Experiments 3 and 5 started in absence of the competing referent. These results together supported the two sources contributing to facilitated maintenance of complex NPs: increased distinctiveness and increased encoding effort.

For internal coherence, the experiments showed that the early maintenance of internally coherent NPs was facilitated. Both Experiments 4a and 4b demonstrated faster RTs during early maintenance for conditions with NPs having similar referents. These results were complemented by the SAN observed in conditions with dissimilar-referent NPs relative to those with similar-referent NPs during early maintenance. Both behavioral and neural evidence pointed to reduced effort to maintain internally coherent target NPs.

Retrieval. There was limited evidence that higher representational complexity or higher internal coherence of the target NP facilitate the retrieval process. Only behavioral experiments with pre-modified complex NPs, Experiments 1a and 1b, showed faster RT during retrieval when the target NP was pre-modified, consistent with the results observed by Hofmeister (2011). The other behavioral experiments and the EEG experiments showed no or weak evidence for the facilitatory effect of retrieving more complex or more coherent target NPs.

5.2 General Discussion and Implications

5.2.1 Working Memory Processes in Language Comprehension

Besides effects of representational complexity and internal coherence on the WM processes, there are broader implications from the experimental results that can contribute to the general understanding of WM processes during language comprehension. Firstly, the experiments showed maintenance site's effects that provided strong support for theories of encoding interference (Nairne, 2002; Oberauer and Kliegl, 2006) and some support for preemptive retrieval. Secondly, they also lent some evidence towards the active maintenance of syntactic features but not semantic features in WM (Wagers and Phillips, 2014). Lastly, the experiments on internal coherence allowed us to speculate about how linguistic referents and feature are organized into a single unit.

Preemptive retrieval and encoding interference. All of the experiments in the dissertation elicited effects during the maintenance site, which can be partially accounted for by the increased distinctiveness between competing referents and reduced vulnerability to interference. However, previous literature has only described interference effects occurring at the retrieval site (Lewis and Vasishth, 2005), not during the maintenance site. Mertzen et al. (2023) observed interference effects prior to retrieval and proposed that they were consequences of preemptive retrieval and/or encoding interference. These processes together can account for the extent of maintenance site's effects observed in our experiments.

Preemptive retrieval is where our language system makes predictions about upcoming linguistic materials and retrieves the target representation before the retrieval site. In our experiments 1a, 1b, 2a and 2b, the facilitatory effect of maintaining more complex representations occurred on the preverbal adverb. It is possible that when participants read the preverbal adverb, they predicted the upcoming verb and retrieved the target NP preemptively. Since complex target NPs are more distinct from competing referents, it is easier to preemptively retrieve them. However, preemptive retrieval alone cannot account for the full extent of the maintenance site's effects. Firstly, these effects did not always occur immediately prior to the retrieval site, like in Experiments 4a, 4b and 5. Secondly, even if the effects occurred prior to the retrieval site, like in Experiments 1a, 1b, 2a and 2b, the effects observed in the maintenance site did not share the same pattern as effects observed in the retrieval site. In most cases, maintenance site's effects were more robust than retrieval site's effects. Lastly, the occurrence of sustained signals from encoding and throughout maintenance in Experiments 3 and 5 indicated that what we observed could not be explained solely by preemptive retrieval. Nevertheless, we cannot reject the possibility of preemptive retrieval partially explaining the results.

On the other hand, our results are consistent with and even provide strong support for the idea of encoding interference (Nairne, 2002; Oberauer and Kliegl, 2006). Many of our effects occurred distally from the retrieval site and thus, could not be explained by *cue overload* (Lewis and Vasishth, 2005). Instead, having a complex target NP with more features can lessen the impact of *feature overwriting* when a new referent is introduced. After feature overwriting, a complex representation would have more intact features left compared to its simple counterpart, resulting in higher activation level of the target NP and thus requiring less maintenance effort. Additionally, in Experiments 1a, 1b, 2a and 2b, the facilitatory effects occurred on words that immediately followed the introduction of a new competing referent, suggesting that the higher representational complexity of the target NP can aid the encoding of these new referents and the maintenance of competing referents.

Active maintenance of features. The results of our experiments also lent some evidence to the active maintenance of syntactic features and not semantic features of representations in WM. A question in the sentence comprehension literature is whether referents are actively maintained in WM or are let decayed and if they are actively maintained, whether the whole referent is or only a part of it. The prevailing view is that syntactic features of linguistic representations are maintained while semantic features are not and thus, decay gradually (Ness and Meltzer-Asscher, 2017,1; Wagers and Phillips, 2014).

Our experiments provided some support for this theory based on the time course differences between the maintenance's site effects. In experiments that manipulated representational complexity, Experiments 1a, 1b, 2a and 2b, maintenance site's effects happened as late as the adverb immediately prior to the retrieval site. Since manipulation of representational complexity changed both syntactic and semantic features of the NP, the differences between the conditions could have lasted longer since the differences in syntactic features were actively maintained. To the contrary, when internal coherence was manipulated in Experiments 4a and 4b, facilitatory effects only lasted two words into the maintenance site. Because manipulation of internal coherence only changed semantic features but not syntactic features, it is likely that the benefit gained from more coherent representations decayed gradually as the semantic features decayed from WM or was relegated to LTM, resulting in effects being absent during late maintenance. Furthermore, the SAN associated with the coherence manipulation in Experiment 5 decreased in magnitude over time, reinforcing the idea that semantic features slowly decay. As a whole, our experiments supported the proposal that syntactic features of representations are actively maintained in WM while semantic features experience gradual decay or are displaced to activated LTM.

Organization of referents and features. The last three experiments which manipulated internal coherence showed that increasing the internal coherence of the target NP is beneficial to all three WM processes, although the facilitation during retrieval is limited. The strongest facilitation effect occurred during late encoding and early maintenance. While these three experiments manipulated the similarity between referents, we did not observe similarity-based interference (Lewis and Vasishth, 2005). Instead, having similar coordinated referents actually facilitated all the WM processes. This is likely the case because these referents were chunked together and as a result, would be retrieved as an ensemble (a coordinated NP).

The facilitatory effects of higher internal coherence could have resulted from more efficient organization of the features from similar referents (Mate and Baqués, 2009), which allows for ease of encoding. It is also possible that in coherent representations, the features are more compatible to each other and prime each other, leading to faster access of the features and decreased encoding time. During maintenance, more efficient organization of fewer features decreases storage cost, resulting in less maintenance effort. Alternatively, when the referents are similar, people are more likely to chunk them together into a constituent. Since it is easier to group together coherent items into a chunk in WM Cowan (2001), it is possible that participants prefer to chunk together similar, coherent referents and to leave dissimilar, incoherent referents as individuated representations. The increased likelihood to chunk similar referents can lead to lower number of maintained representations on average and thus, lower maintenance cost.

5.2.2 Persistent Neural Activity in Language Comprehension

Another broader implication of the dissertation is to contribute to the general understanding of the SAN and sustained EEG signals in general during sentence comprehension. First, our results suggested that the SAN reflects increased WM load but might also be linked to predictive processes. Secondly, our observations about different appearances of the SAN in the literature generated speculations about the distinctions and a possible link to activity-silent WM representations. Lastly, our observations of the SAN suggest that it is sensitive to the content of the WM representations, contrary to other sustained EEG signals such as the CDA.

Functional interpretation of the SAN and other sustained signals. One of the goals of the dissertation is to help inform the functional significance of the SAN and other sustained signals in sentence comprehension. In Experiment 3, we detected a global negativity during encoding and maintenance in the simple NP conditions relative to the complex NP conditions. We also found a SAN during maintenance in conditions with SRC relative to those with ORC and a SAN during retrieval in conditions with ORC relative to those with SRC. In Experiment 5, we detected a global negativity during encoding and maintenance in the simple NP conditions. Additionally, having a simpler or incoherent target NP both elicited the SAN from the last word of the encoding site to the early maintenance site, relative to having a more complex or coherent NP, respectively.

The significance of the global negativities in Experiments 3 and 5 is hard to decipher since they can be artifacts of pre-baselining due to different pre-baseline materials. However, their directionality was in line with other sustained signals occurring during WM maintenance. Like the CDA, NSW and SAN, the more WM-intensive conditions with simple target NPs elicited more global negativity than the less WM-intensive conditions with complex target NPs.

Besides the global negativities, the other instances of persistent neural activity we observed were all manifestations of the SAN. In our experiments, the SAN reflected difficulty in processing during the maintenance period. This difficulty could arise from many sources, such as increased memory demands from reading ORCs, the difficulty associated with reading two consecutive adverbs or lower activation level associated with maintaining a simple or incoherent target NP.

While all these sources presented processing challenges, not all of them increased the number of representations in WM. In particular, the SAN in SRCs compared to ORCs in Experiment 3, which can be attributed to the challenges of reading an infrequent sequence of two adverbs, cannot be explained by an increase in the number of representations. At this point, sentences with ORCs had one more referent than sentences with SRCs, the RC subject. This suggested that the SAN might be sensitive to other factors, such as participants' expectation of the upcoming linguistic material.

In the sentence comprehension literature, expectation-based theories highlighted that unmet expectations can generate processing difficulty (Hale, 2001; Levy, 2008). Comprehenders made predictions about upcoming linguistic input, which are partially built upon their previous experience. Violations of these predictions can increase processing cost. Since a sequence of two adverbs like *today harshly* is unusual, this could have led to increased processing cost. There has also been other evidence that expectation can influence behavior of the SAN. Yano and Koizumi (2018) and Yano and Koizumi (2021) reported that while non-canonical scrambled word order in Japanese could elicit the SAN relative to canonical unscrambled word order, the SAN disappeared when the non-canonical word order was licensed by the right context. In these cases, prior context led readers to assign higher probabilities to the scrambled word order, thus incurring less processing cost. It is still challenging to bring together load-based and expectation-based explanations of the SAN. It is possible that the SAN reflects shared resources for both storage and processing (Gibson, 2000). Increased cost to either storage or processing can elicit the SAN.

Different appearances of the SAN. Another noteworthy but speculative observation from our results is that the various SANs reported in the literature and observed by us appear to come in two kinds based on their visual appearance. Visually speaking, the SANs, such as those observed by King and Kutas (1995) (Fig 5.1) and Schrum and Sprouse (2022), did not experience a significant magnitude decrease throughout their occurrence and lasted past the final word of the sentence. I will label these as continuous SANs. To the contrary, the SANs observed in Experiment 5, as well as those reported by Fiebach et al. (2002); Lau (2018); Phillips et al. (2005), showed a repeated pattern where the SAN decreased in magnitude at the onset of an incoming word and increased in magnitude around its offset. I will label these as discontinuous SANs.

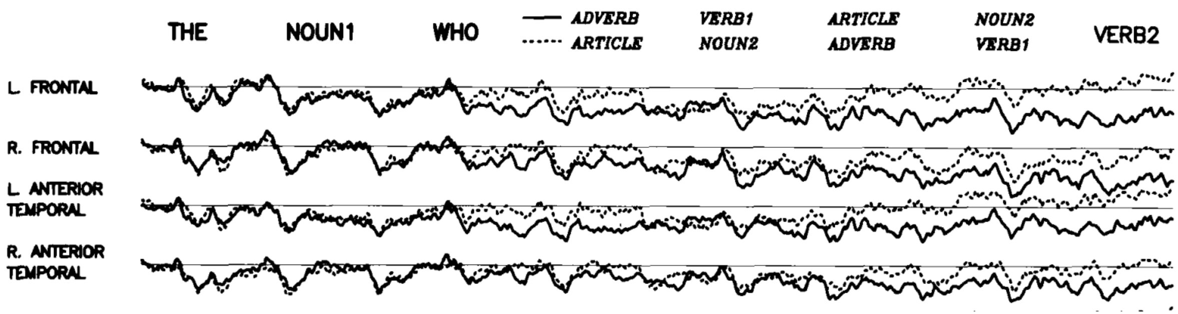

Figure 5.1: Continuous SAN. SOURCE: King and Kutas (1995). Reproduced with permission from *MIT Press*.

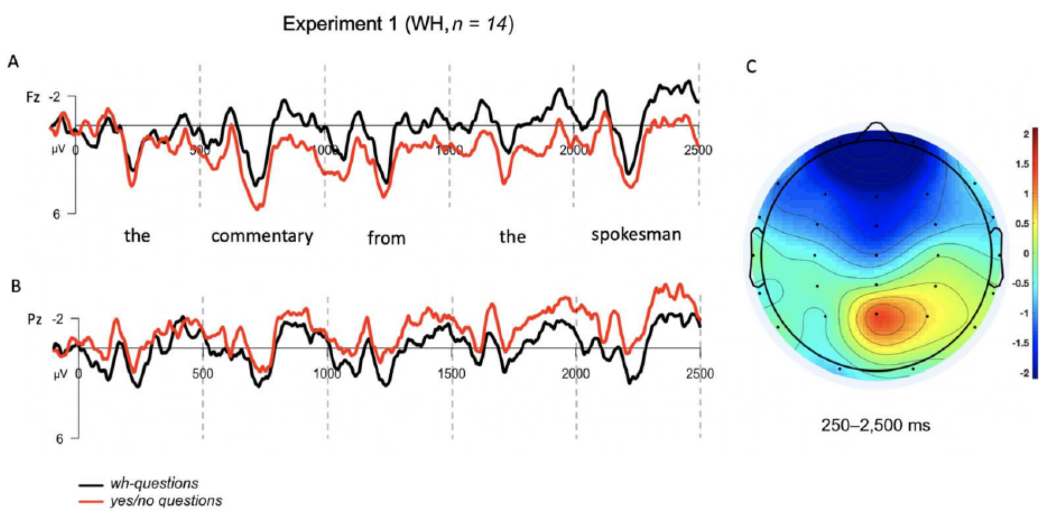

Figure 5.2: Discontinuous SAN. SOURCE: Cruz Heredia et al. (2022). Reproduced with permission from *MIT Press*.

It is unclear what factors produced this distinction. Notably, however, King and Kutas (1995) compared different words belonging to different syntactic categories at the same region (adverb in SRC with article in ORC, verb in SRC with noun in ORC,...), which could have magnified the SAN and resulted in a more continuous appearance. Nevertheless, Schrum and Sprouse (2022) observed the SAN in regions with exactly the same linguistic material.

What can account for the discontinuous SANs' cyclical pattern of increasing and decreasing magnitude? Previous studies of visual WM have observed that disruption during the task can decrease the magnitude of sustained signals like the CDA, the NSW and lateralized alpha power (Berggren and Eimer, 2016; Hakim et al., 2020; Kreither et al., 2022; Yu et al., 2023). Hakim et al. (2020) showed that lateralized alpha power was recovered later in the trial after its disappearance. Kreither et al. (2022) demonstrated that CDA disruptions did not lead to a significant decrease in WM task performance. There are several possible explanations for this phenomenon: the information in WM is converted to activity-silent representations (Kreither et al., 2022), is displaced into activated LTM (Foster et al., 2019) or is there but is not pronounced enough to be picked up by statistical analysis. Whatever the reason is, the cyclical pattern observed in the discontinuous SANs might reflect the same process. It might be the case that with every incoming word, the representations in WM are displaced temporarily to make space for the new word and then recovered once the word has been integrated into the syntactic structure. This observation showed that the SAN (and possibly other sustained signals) does not necessarily manifest as a continuous signal but can also appear as a signal punctuated by disruptions with each incoming word.

Content sensitivity of the SAN. Our EEG experiments also demonstrated that the SAN might be sensitive to both the number of representations maintained in WM and their content. While its magnitude can be modulated by the number of linguistic representations stored in WM (such as between ORCs and SRCs), it is also modulated by representational complexity and coherence between features within a representation, which reflect the content of the representations themselves, not the number of maintained representations. However, unlike the SAN, the CDA and the NSW - ERP indices associated with visual WM load - are only sensitive to the number of representations and not their content (Diaz et al., 2021). In fact, for visual WM, there are two sustained neural codes (Emrich et al., 2013): a frontoparietal code that is load-sensitive and a code in early visual regions that is content-sensitive. It is possible that due to the highly hierarchical nature of language, linguistic WM maintains both load information, as well as the content (or part of the content) of the representations. Further investigations are necessary to comprehend how linguistic units are maintained in WM neurally.

5.3 Limitations and Future Directions

A key limitation of the EEG studies was that due to the different pre-baseline materials, the global negativities which appeared in Experiments 3 and 5 were difficult to interpret because it is unclear whether they reflected real differences between the conditions or were artifacts of the pre-baselining process. We note that this is a challenge that is not unique to our experiments but also to previous studies that elicited the SAN such as Phillips et al. (2005). Since we wanted to manipulate representational complexity based on pre-modification and coordination, it was necessary to insert additional words in the encoding site for the conditions with complex target NPs. In a future attempt, we can use semantically complex NPs (for example, *the nurse*) that do not have more words than their simple counterparts (*the person*).

Another limitation of the EEG results presented here is that there remains many other dimensions to the data that can be explored beyond just sustained EEG differences. Sustained EEG indices and oscillatory power have both been shown to both index WM maintenance while having different behavior and are sensitive to different properties of the memorandum (Diaz et al., 2021; Hakim et al., 2020; Unsworth et al., 2015). We can continue explore our data through analyzing the oscillatory pattern, which might reveal differences that are elusive to our current analysis technique. Previous attempts of analyzing oscillatory patterns in sentence comprehension have yielded interesting relationship between oscillations at different frequencies and different linguistic levels (Ding et al., 2016). Thyer et al. (2022) has also shown that multivariate analysis of scalp EEG topography could track number of items stored in WM and not the content of those items. Application of multivariate analysis on our EEG data might yield load-sensitive and content-insensitive patterns that differ from the SANs that we observed, which are content-sensitive.

Lastly, there remains many questions that we want to continue to explore with regards to the WM architecture in sentence comprehension. So far, all of our experiments have relied on manipulating the content of the target NPs. If a process can help facilitate maintenance by keeping the competing referents distinctive, making an intervening NP more complex should decrease maintenance effort since less effort is required to keep it distinct from the target NP. One of our proximal goals is to test this hypothesis by varying the complexity of non-target NPs and observing the effect of the manipulation on the various WM processes.

Another finding we want to explore further is how the content of linguistic referents is represented in WM. We have suggested that the SAN might be sensitive to both the number of representations and their content. However, WM research in the visual and verbal domains suggested that representations in WM are stored as content-free "pointers" (Huang and Awh, 2018; Kahneman et al., 1992; Thyer et al., 2022). Huang and Awh (2018) showed that when visual or verbal objects were chunked together into a representation, it took a longer time to retrieve individual components compared to when they were not chunked. This was because extra effort was needed to retrieve the content from LTM. We want to adapt this paradigm in the context of sentence comprehension. An ideal test case would involve the retrieval of individual referents from coordinated NPs. Comparison between the effort taken to retrieve a referent from an uncoordinated NP and a coordinated NP can help us understand better the nature of WM representations in sentence comprehension.

APPENDIX A

RAW READING TIME PLOTS FOR SELF-PACED READING EXPERIMENTS

A.1 Experiment 1a

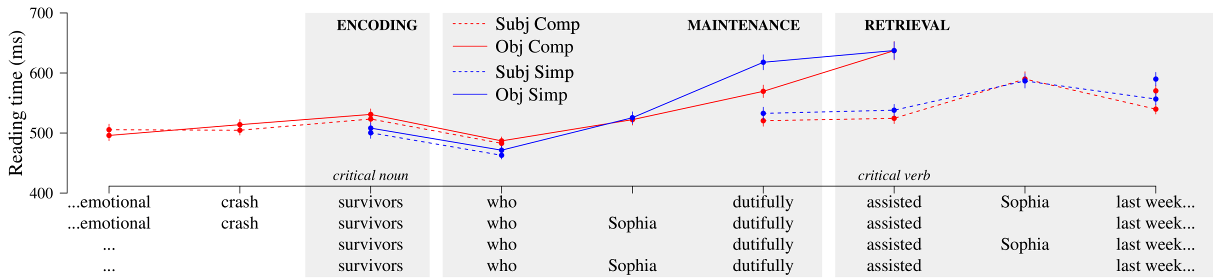

Figure A.1: Experiment 1a: Raw reading times

A.2 Experiment 1b

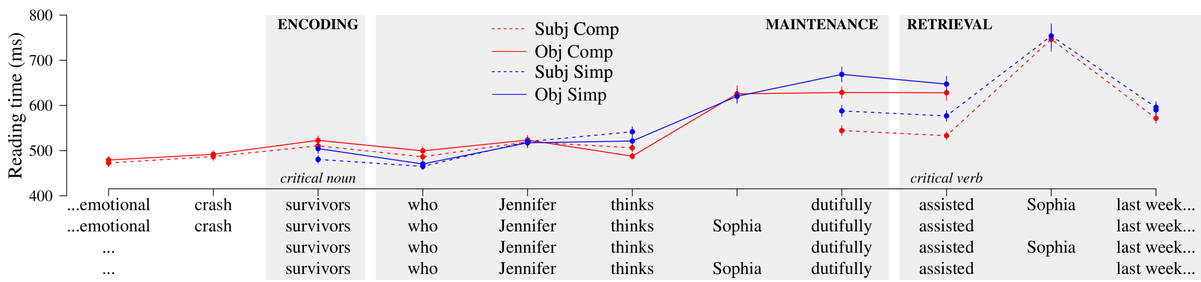

Figure A.2: Experiment 1b: Raw reading times

A.3 Experiment 2a

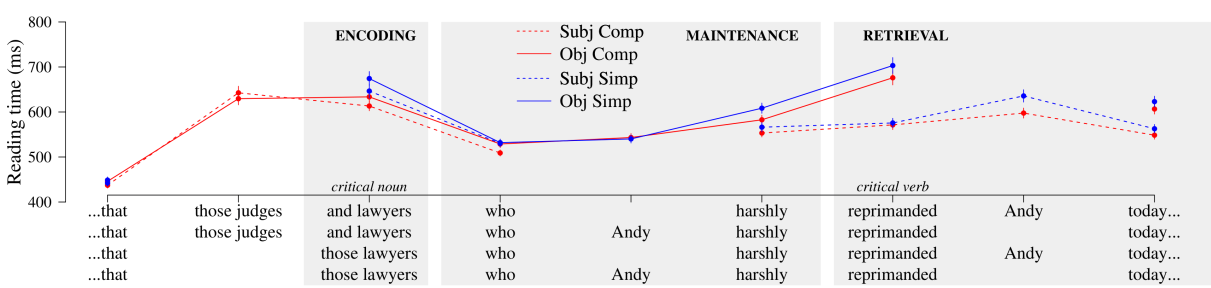

Figure A.3: Experiment 2a: Raw reading times

A.4 Experiment 2b

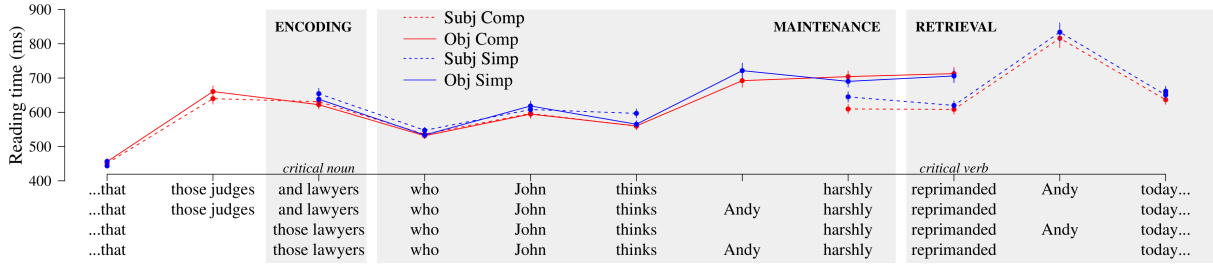

Figure A.4: Experiment 2b: Raw reading times

A.5 Experiment 4a

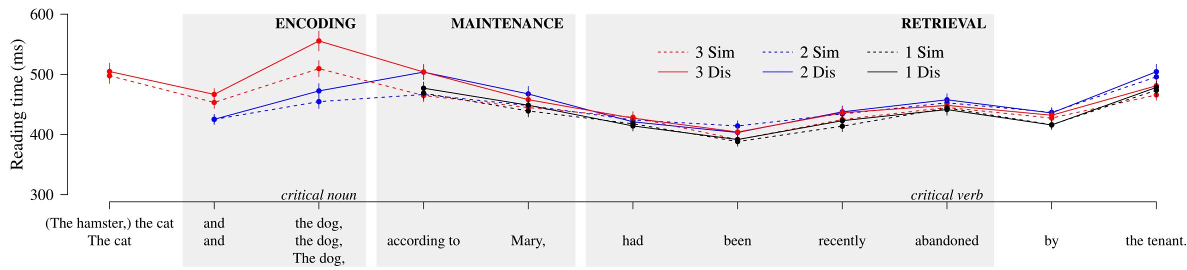

Figure A.5: Experiment 4a: Raw reading times

A.6 Experiment 4b

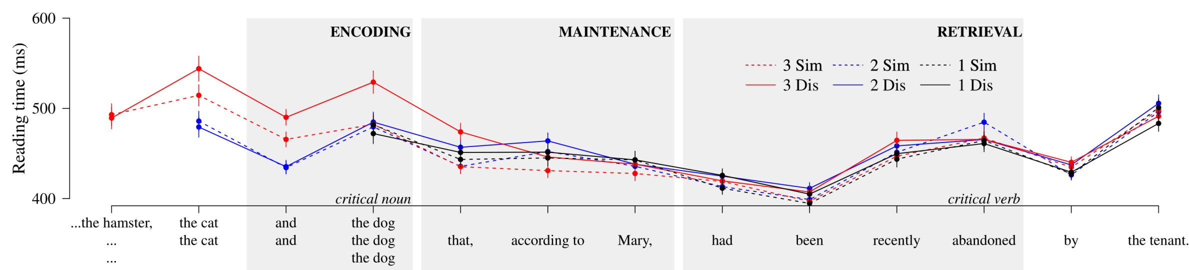

Figure A.6: Experiment 4b: Raw reading times

APPENDIX B

STIMULI SENTENCES FOR EXPERIMENTS 1A/1B

Sentences from Experiment 1a do not contain the underlined additional embedded layer (*John thinks* in sentence 1). Sentences from Experiment 1b contain the underlined materials. All experimental items presented below have SRCs and complex filler NPs. To form conditions with simple filler NPs, remove the two pre-modifiers in front of the matrix subject noun which are in parentheses (*federal prison* in sentence 1). To form conditions with ORCs, move the RC object proper noun (*Andy* in sentence 1) in front of the preverbal adverb (*harshly*). For example, sentence 1 should be ... *Andy harshly reprimanded* ... instead.

B.1 Experimental Items

1. Those (federal prison) wardens who John thinks harshly reprimanded Andy today admitted the error.

2. Those (condemned political) prisoners who Caroline says accidentally angered Amy yesterday escaped the compound.

3. Those (experienced stage) choreographers who Tim believes warmly encouraged Ben last night created the routine.

4. Those (sleazy car) salespeople who Jude claims deviously tricked Luke this morning left the company.

5. Those (graceful ballet) dancers who George says wholeheartedly praised Kevin this afternoon won the competition.

6. Those (wounded American) soldiers who Sarah believes graciously commended Leah this evening organized the march.

7. Those (hilarious stand-up) comedians who Ali claims thoroughly entertained Jack tonight sold the tickets.

8. Those (emotional crash) survivors who Jennifer thinks dutifully assisted Sophia last week joined the meeting.
9. Those (peaceful Buddhist) monks who Nico believes loyally aided Matthew on Monday repeated the chants.
10. Those (conservative US) senators who Jerry claims carefully interrogated Tom on Tuesday wrote the bill.
11. Those (victorious four-star) generals who Ruth thinks profusely thanked Stephanie on Wednesday signed the treaty.
12. Those (struggling rock) musicians who Hallie says reluctantly helped Becky on Thursday performed the song.
13. Those (alleged bombing) accomplices who Jane claims vehemently accused Ellen on Friday accepted the charges.
14. Those (undercover federal) agents who Elsa thinks secretly confronted Anna on Saturday refused the deal.
15. Those (celebrity hair) stylists who Barack says hesitantly hired Teddy on Sunday closed the store.
16. Those (renowned fashion) designers who Kennedy believes snidely belittled Alex today designed the costume.
17. Those (senior electrical) engineers who Mark thinks affectionately mentored Zack yesterday fixed the machine.
18. Those (famous young) actors who Zoey says caringly guided Madeline last night read the script.
19. Those (hard-nosed newspaper) reporters who Michelle believes ruthlessly drilled Sasha this morning attended the game.

20. Those (wealthy teen) celebrities who Elisa claims continuously questioned Fiona this afternoon wasted the money.

21. Those (corrupt homicide) cops who Bilbo says aggressively accosted Ivan this evening issued the statement.

22. Those (ruthless military) dictators who Ewan believes intentionally ignored Isaac tonight skipped the hearing.

23. Those (hardworking factory) employees who Jackie claims brutally berated Alma last week finished the project.

24. Those (cable news) analysts who Corey thinks purposely neglected Ian on Monday published the article.

25. Those (liberal mayoral) candidates who Savannah believes cordially welcomed Rosie on Tuesday threw the party.

26. Those (rich European) tourists who Catherine claims blindly followed Isabella on Wednesday took the photographs.

27. Those (powerful business) executives who Britta thinks diligently instructed Jessica on Thursday hosted the lunch.

28. Those (influential legal) advisors who Annie says meticulously counseled Tess on Friday donated the fund.

29. Those (angry taxi) drivers who Zane claims viciously attacked Keith on Saturday fled the scene.

30. Those (helpful new) assistants who Troy thinks calmly ushered Simon on Sunday decorated the office.

31. Those (dying elderly) patients who Maria says lovingly comforted Elena today received the vaccine.

32. Those (trained hospice) nurses who Laura believes painstakingly watched Aliza yesterday reached the hospital.

B.2 Filler Items

1. It was those consumers who Connor thinks begrudgingly contacted Jack last night.
2. It was those landlords who Frankie says generously paid Ariana this morning.
3. It was those builders who Chad believes contentedly contracted Grant this afternoon.
4. It was those administrators who Serena claims devotedly obeyed Anne this evening.
5. It was those marathoners who Venus says Juliette excessively trained tonight.
6. It was those shareholders who Jeff believes William steadfastly trusted last week.
7. It was those archaeologists who Grace claims Abby eagerly protested on Monday.
8. It was those ministers who Ava thinks Katie constantly distracted on Tuesday.
9. Carlos believes that the fact that those writers beautifully described Sergio this afternoon was unprecedented.
10. Joe claims that the fact that those publishers positively endorsed Thomas this evening was rare.
11. Diego thinks that the fact that those chemists arbitrarily selected Luis tonight was unusual.
12. Larry says that the fact that those planners carelessly mocked Lynn last week was normal.
13. Derrick claims that the fact that Jacob negatively rated those recruiters on Monday was surprising.

14. Frank thinks that the fact that Terry surreptitiously observed those investigators on Tuesday was astonishing.
15. Charlotte says that the fact that Kacey attentively shadowed those clinicians on Wednesday was impressive.
16. Evelyn believes that the fact that Blythe confidently taught those artisans on Thursday was bizarre.
17. It was mind-boggling that those florists nervously notified Nick on Saturday.
18. It was uncanny that those orthodontists unconditionally forgave Bennet on Sunday.
19. It was acceptable that those midwives fearfully alerted Bill today.
20. It was unacceptable that those marines fiercely assaulted Lucas yesterday.
21. It was abnormal that Sadie unethically misinformed those entrepreneurs last night.
22. It was unbelievable that Xavier unconsciously misled those captains this morning.
23. It was incredible that Chad foolishly indulged those cooks this afternoon.
24. It was mind-blowing that Leland severely underpaid those pedicurists this evening.
25. The newspaper confirmed that those senators mistakenly exposed Tabitha last week.
26. The article acknowledged that those treasurers incorrectly chastised Preston on Monday.
27. The show explained that those chancellors publicly honored Sienna on Tuesday.
28. The book demonstrated that those landscapers overtly defended Justin on Wednesday.
29. The journal established that Selena sincerely complimented those superintendents on Thursday.
30. The evidence proved that Taylor maliciously antagonized those cobblers on Friday.

31. The company recognized that Ethan arrogantly corrected those interpreters on Saturday.

32. The documentary revealed that Olivia loudly lauded those barbers on Sunday.

APPENDIX C

STIMULI SENTENCES FOR EXPERIMENTS 2A/2B

Sentences from Experiment 2a do not contain the underlined additional embedded layer (*John thinks* in sentence 1). Sentences from Experiment 2b contain the underlined materials. All experimental items presented below have SRCs and complex filler NPs. To form conditions with simple filler NPs, remove the first noun of the matrix subject NP and *and* which are in parentheses (*lawyers and* in sentence 1). To form conditions with ORCs, move the RC object proper noun (*Andy* in sentence 1) in front of the preverbal adverb (*harshly*). For example, sentence 1 should be ... *Andy harshly reprimanded* ... instead.

C.1 Experimental Items

1. It seems that those (lawyers and) judges who <u>John thinks</u> harshly reprimanded Andy today admitted the error.

2. It appears that those (designers and) decorators who <u>Caroline says</u> inadvertently angered Amy yesterday furnished the mansion.

3. It seems that those (biologists and) researchers who <u>Tim believes</u> warmly encouraged Ben last night studied the disease.

4. It appears that those (publicists and) managers who <u>Jude claims</u> diligently represented Luke this morning answered the letter.

5. It seems that those (singers and) dancers who <u>George says</u> mercilessly derided Kevin this afternoon sold the tickets.

6. It appears that those (clerks and) coordinators who <u>Sarah believes</u> rudely criticized Leah this evening organized the calendar.

7. It seems that those (cashiers and) bussers who <u>Ali claims</u> adequately served Jack tonight prepared the food.

8. It appears that those (dieticians and) technicians who Jennifer thinks dutifully assisted Sophia last week administered the test.
9. It seems that those (photographers and) journalists who Nico believes loyally aided Matthew on Monday broke the story.
10. It appears that those (composers and) pianists who Jerry claims lovingly comforted Tom on Tuesday wrote the sonata.
11. It seems that those (policemen and) firemen who Ruth thinks swiftly called Stephanie on Wednesday rescued the dog.
12. It appears that those (psychologists and) psychiatrists who Hallie says seriously doubted Becky on Thursday prescribed the medication.
13. It seems that those (painters and) framers who Jane claims emphatically denounced Ellen on Friday completed the artwork.
14. It appears that those (electricians and) plumbers who Elsa thinks reluctantly helped Anna on Saturday collected the payment.
15. It seems that those (tailors and) seamstresses who Barack says halfheartedly hired Teddy on Sunday created the clothing.
16. It appears that those (engineers and) contractors who Kennedy believes eagerly introduced Alex today renovated the hotel.
17. It seems that those (teachers and) professors who Mark thinks affectionately mentored Zack yesterday started the lesson.
18. It appears that those (veterinarians and) zoologists who Zoey says quickly advised Madeline last night treated the animal.
19. It seems that those (coaches and) directors who Michelle believes ruthlessly drilled Sasha this morning attended the game.

20. It appears that those (bankers and) tellers who Elisa claims thoroughly questioned Fiona this afternoon counted the cash.

21. It seems that those (librarians and) volunteers who Bilbo says enthusiastically guided Ivan this evening sorted the books.

22. It appears that those (witnesses and) defendants who Ewan believes frighteningly interrogated Isaac tonight skipped the hearing.

23. It seems that those (clients and) patrons who Jackie claims hurriedly command)ed Amy last week finished the project.

24. It appears that those (repairmen and) mechanics who Corey thinks consciously neglected Ian on Monday repaired the bike.

25. It seems that those (socialites and) influencers who Savannah believes snobbishly excluded Rosie on Tuesday threw the party.

26. It appears that those (comedians and) performers who Catherine claims barely recognized Isabella on Wednesday coordinated the show.

27. It seems that those (secretaries and) bosses who Britta thinks carefully counseled Jessica on Thursday hosted the lunch.

28. It appears that those (doctors and) nurses who Annie says previously consulted Tess on Friday diagnosed the flu.

29. It seems that those (artists and) sculptors who Zane claims cordially welcomed Keith on Saturday made the piece.

30. It appears that those (conductors and) accompanists who Troy thinks fully prepared Simon on Sunday practiced the performance.

31. It seems that those (medics and) EMTs who Maria says successfully relieved Elena today reached the hospital.

32. It appears that those (therapists and) counselors who Laura believes compassionately consoled Aliza yesterday alleviated the problem.

C.2 Filler Items

1. It was those consumers who Connor thinks begrudgingly contacted Jack last night.
2. It was those landlords who Frankie says generously paid Ariana this morning.
3. It was those builders who Chad believes contentedly contracted Grant this afternoon.
4. It was those administrators who Serena claims devotedly obeyed Anne this evening.
5. It was those marathoners who Venus says Juliette excessively trained tonight.
6. It was those shareholders who Jeff believes William steadfastly trusted last week.
7. It was those archaeologists who Grace claims Abby blindly followed on Monday.
8. It was those ministers who Ava thinks Katie constantly distracted on Tuesday.
9. Carlos believes that the fact that those writers beautifully described Sergio this afternoon was unprecedented.
10. Joe claims that the fact that those publishers positively endorsed Thomas this evening was rare.
11. Diego thinks that the fact that those chemists arbitrarily selected Luis tonight was unusual.
12. Larry says that the fact that those planners carelessly mocked Lynn last week was normal.
13. Derrick claims that the fact that Jacob negatively rated those recruiters on Monday was surprising.

14. Frank thinks that the fact that Terry secretly observed those investigators on Tuesday was astonishing.

15. Charlotte says that the fact that Kacey attentively shadowed those clinicians on Wednesday was impressive.

16. Evelyn believes that the fact that Blythe confidently instructed those artisans on Thursday was bizarre.

17. It was mind-boggling that those florists nervously notified Nick on Saturday.

18. It was uncanny that those orthodontists graciously forgave Bennet on Sunday.

19. It was acceptable that those midwives fearfully alerted Bill today.

20. It was unacceptable that those marines fiercely attacked Lucas yesterday.

21. It was abnormal that Sadie unethically misinformed those entrepreneurs last night.

22. It was unbelievable that Xavier accidentally misled those captains this morning.

23. It was incredible that Chad foolishly indulged those cooks this afternoon.

24. It was mind-blowing that Leland severely underpaid those pedicurists this evening.

25. The newspaper confirmed that those senators mistakenly exposed Tabitha last week.

26. The article acknowledged that those treasurers incorrectly chastised Preston on Monday.

27. The show explained that those chancellors publicly honored Sienna on Tuesday.

28. The book demonstrated that those landscapers overtly defended Justin on Wednesday.

29. The journal established that Selena sincerely complimented those superintendents on Thursday.

30. The evidence proved that Taylor maliciously antagonized those cobblers on Friday.

31. The company recognized that Ethan arrogantly corrected those interpreters on Saturday.

32. The documentary revealed that Olivia loudly lauded those barbers on Sunday.

APPENDIX D
STIMULI SENTENCES FOR EXPERIMENT 3

All experimental items presented below have SRCs and complex filler NPs. To form conditions with simple filler NPs, remove the two pre-modifiers in front of the matrix subject noun which are in parentheses (*merciless prison* in sentence 1). To form conditions with ORCs, switch positions between the RC object proper noun (*Andy* in sentence 1) and the adverb of time (*today*). For example, sentence 1 should be *... Andy harshly reprimanded today ...* instead.

D.1 Experimental Items

1. It appears that the (merciless prison) warden who today harshly reprimanded Andy admitted the error.

2. It appears that the (condemned political) prisoner who yesterday accidentally angered Amy escaped the compound.

3. It appears that the (experienced stage) choreographer who tonight warmly encouraged Ben created the routine.

4. It appears that the (sleazy car) salesperson who last night deviously tricked Luke quit the company.

5. It appears that the (graceful ballet) dancer who this week wholeheartedly praised Kevin won the competition.

6. It appears that the (wounded American) soldier who last week graciously commended Leah organized the march.

7. It appears that the (hilarious stand-up) comedian who on Monday thoroughly entertained Jack sold the tickets.

8. It appears that the (emotional crash) survivor who on Tuesday dutifully assisted Sophia joined the meeting.

9. It appears that the (peaceful Buddhist) monk who on Wednesday loyally aided Matthew repeated the chants.

10. It appears that the (conservative US) senator who on Thursday carefully interrogated Tom wrote the bill.

11. It appears that the (victorious four-star) general who on Friday profusely thanked Stephanie signed the treaty.

12. It appears that the (struggling rock) musician who on Saturday greatly helped Becky performed the song.

13. It appears that the (alleged bombing) accomplice who on Sunday vehemently accused Ellen denied the charges.

14. It appears that the (undercover federal) agent who today secretly confronted Anna refused the deal.

15. It appears that the (celebrity hair) stylist who yesterday hesitantly hired Teddy closed the store.

16. It appears that the (renowned fashion) designer who tonight snidely belittled Alex designed the costume.

17. It appears that the (senior electrical) engineer who last night affectionately mentored Zack fixed the machine.

18. It appears that the (famous young) actor who this week caringly guided Madeline read the script.

19. It appears that the (hard-nosed newspaper) journalist who last week ruthlessly drilled Sasha attended the game.

20. It appears that the (spoiled teen) celebrity who on Monday continuously questioned Fiona wasted the opportunity.

21. It appears that the (corrupt homicide) cop who on Tuesday aggressively accosted Ivan issued the statement.

22. It appears that the (ruthless military) dictator who on Wednesday intentionally ignored Issac skipped the hearing.

23. It appears that the (hardworking factory) employee who on Thursday brutally berated Alma finished the project.

24. It appears that the (scrupulous financial) analyst who on Friday purposely neglected Ian crunched the numbers.

25. It appears that the (liberal mayoral) candidate who on Saturday cordially welcomed Rosie threw the party.

26. It appears that the (rich European) tourist who on Sunday blindly followed Isabella took the photographs.

27. It appears that the (powerful business) executive who today diligently instructed Jessica hosted the lunch.

28. It appears that the (influential legal) advisor who yesterday meticulously counseled Tess donated the money.

29. It appears that the (angry taxi) driver who tonight viciously attacked Keith fled the scene.

30. It appears that the (helpful new) assistant who last night calmly pacified Simon decorated the office.

31. It appears that the (dying elderly) patient who this week lovingly comforted Elena received the vaccine.

32. It appears that the (trained hospice) nurse who last week painstakingly watched Aliza reached the hospital.

33. It appears that the (passionate defense) lawyer who on Monday patiently persuaded Justin submitted the evidence.

34. It appears that the (stern district) judge who on Tuesday coolly advised Daniel entered the courtroom.

35. It appears that the (meticulous interior) decorator who on Wednesday halfheartedly recommended Fiora arranged the furniture.

36. It appears that the (eminent marine) biologist who on Thursday grudgingly acknowledged Lin got the grant.

37. It appears that the (distinguished cancer) researcher who on Friday successfully reassured Ross discovered the cell.

38. It appears that the (disgruntled unpaid) intern who on Saturday really astonished Rachel rejected the offer.

39. It appears that the (strict interim) manager who on Sunday deliberately avoided Monica established the committee.

40. It appears that the (talented blues) singer who today unwittingly chose Chandler cancelled the program.

41. It appears that the (diligent office) clerk who yesterday noisily cheered Joey completed the form.

42. It appears that the (sloppy event) coordinator who tonight politely consulted Phoebe counted the attendees.

43. It appears that the (nosy store) cashier who last night nearly convinced Shen faced the consequences.

44. It appears that the (junior lab) technician who this week arrogantly criticized Janice halted the experiment.

45. It appears that the (professional landscape) photographer who last week stupidly dared John discussed the picture.

46. It appears that the (creative classical) composer who on Monday almost doubted Lenore produced the masterpiece.

47. It appears that the (gifted concert) pianist who on Tuesday clearly discouraged Jason played the piece.

48. It appears that the (amateur acoustic) guitarist who on Wednesday obviously enjoyed Laura practiced the tune.

49. It appears that the (decorated city) policewoman who on Thursday forcefully examined Rodrigo solved the mystery.

50. It appears that the (dedicated volunteer) fireman who on Friday luckily found Eszter saved the cat.

51. It appears that the (significant cognitive) psychologist who on Saturday absentmindedly forgot Mike measured the behavior.

52. It appears that the (accredited clinical) psychiatrist who on Sunday willingly forgave Aurora scheduled the appointment.

53. It appears that the (brilliant watercolor) painter who today severely hurt Han admired the painting.

54. It appears that the (careful residential) electrician who yesterday quickly identified Ilya repaired the socket.

55. It appears that the (reliable maintenance) plumber who tonight surprisingly impressed Rich unclogged the pipe.

56. It appears that the (stylish French) tailor who last night privately informed Grace made the dress.

57. It appears that the (fashionable Italian) seamstress who this week enthusiastically introduced Emre sewed the curtains.

58. It appears that the (expensive general) contractor who last week unnecessarily involved Joan sealed the contract.

59. It appears that the (thoughtful school) teacher who on Monday gently led Jaren cleaned the board.

60. It appears that the (tenured university) professor who on Tuesday briefly mentioned Harold taught the class.

61. It appears that the (gentle wildlife) veterinarian who on Wednesday reluctantly billed Harry euthanized the animal.

62. It appears that the (knowledgeable English) zoologist who on Thursday unfortunately missed Emma explored the forest.

63. It appears that the (observant tennis) coach who on Friday attentively observed Fred ruined the racket.

64. It appears that the (quirky movie) director who on Saturday indirectly complimented George directed the film.

65. It appears that the (greedy investment) banker who on Sunday lethally punched Mia embezzled the fund.

66. It appears that the (old academic) librarian who today savagely kicked Jules hid the dictionary.

67. It appears that the (eager campaign) volunteer who yesterday forcibly pulled Gabby abandoned the cause.

68. It appears that the (nervous key) witness who tonight surreptitiously pursued Omer forfeited the weapon.

69. It appears that the (persuasive criminal) defendant who last night maliciously pushed Carlos lost the case.

70. It appears that the (demanding regular) client who this week repeatedly bugged Caitlyn typed the review.

71. It appears that the (loud museum) patron who last week angrily called Garen violated the law.

72. It appears that the (resourceful fridge) repairman who on Monday rudely infuriated Hans replaced the screw.

73. It appears that the (skilled truck) mechanic who on Tuesday happily drove Jonie patched the tire.

74. It appears that the (wealthy British) socialite who on Wednesday momentarily saw Paul squandered the inheritance.

75. It appears that the (vain Instagram) influencer who on Thursday boastfully emailed Minnie posted the selfie.

76. It appears that the (captivating circus) performer who on Friday gladly replaced Kate perfected the trick.

77. It appears that the (loyal personal) secretary who on Saturday sympathetically texted Charles brought the coffee.

78. It appears that the (controlling company) boss who on Sunday sheepishly nudged Kwame ate the sandwich.

79. It appears that the (astute throat) doctor who today fatally shot Lydia patented the cure.

80. It appears that the (rising pop) artist who yesterday fearlessly robbed Albert auctioned the portrait.

81. It appears that the (thorough marble) sculptor who tonight abruptly stopped Nelson chiseled the bust.

82. It appears that the (watchful train) conductor who last night apparently threatened Andrew caused the crash.

83. It appears that the (brave combat) medic who this week voluntarily taught Mark left the army.

84. It appears that the (friendly massage) therapist who last week fiercely grabbed Vincent broke the table.

85. It appears that the (encouraging college) counselor who on Monday exceedingly surprised Flora sued the institution.

86. It appears that the (scary bank) robber who on Tuesday stealthily warned Derek stole the jewelry.

87. It appears that the (sophisticated art) thief who on Wednesday supposedly intimidated Olivia dropped the gun.

88. It appears that the (cruel serial) killer who on Thursday unexpectedly tolerated Ava committed the murder.

89. It appears that the (budding fiction) writer who on Friday verbally abused Mia printed the novel.

90. It appears that the (active global) citizen who on Saturday eagerly accepted Charlotte traveled the globe.

91. It appears that the (mindful world) traveler who on Sunday readily accomodated Amelia missed the flight.

92. It appears that the (alert security) guard who today utterly annoyed Harper detected the bombs.

93. It appears that the (agile soccer) player who yesterday impolitely answered Evelyn scored the goal.

94. It appears that the (fit Kenyan) athlete who tonight unknowingly antagonized Abigail beat the record.

95. It appears that the (driven Korean) sportswoman who last night delightfully applauded Alexis waved the flag.

96. It appears that the (eloquent crime) podcaster who this week horrifically assaulted Emily started the trend.

97. It appears that the (systematic software) programmer who last week sharply attacked Ella programmed the tool.

98. It appears that the (extraordinary chess) prodigy who on Monday sarcastically bad-mouthed Camille sacrificed the queen.

99. It appears that the (imaginative fantasy) author who on Tuesday entirely baffled Aria read the passage.

100. It appears that the (overworked tax) accountant who on Wednesday silently begged Penelope calculated the invoices.

101. It appears that the (licensed commercial) architect who on Thursday shamelessly betrayed Avery drafted the blueprint.

102. It appears that the (aggressive divorce) attorney who on Friday partially blamed Mila bought the condo.

103. It appears that the (chatty local) barber who on Saturday publicly bribed Scarlett reopened the shop.

104. It appears that the (snobbish club) bartender who on Sunday unmercifully bullied Layla mixed the cocktail.

105. It appears that the (sweaty home) builder who today tirelessly chased Chloe laid the groundwork.

106. It appears that the (ambitious corporate) businesswoman who yesterday heartily congratulated Victoria purchased the stock.

107. It appears that the (paranoid army) colonel who tonight promptly contacted Eleanor shot the plane.

108. It appears that the (taciturn ship) carpenter who last night hastily corrected Nora sawed the tree.

109. It appears that the (jolly dessert) chef who this week vigorously defended Riley frosted the cake.

110. It appears that the (amicable family) dentist who last week pleasantly greeted Zoey removed the cavity.

111. It appears that the (weary game) developer who on Monday vividly described Hannah licensed the product.

112. It appears that the (fraudulent vegan) dietician who on Tuesday curtly dismissed Hazel counterfeited the pills.

113. It appears that the (unethical behavioral) economist who on Wednesday willfully disobeyed Violet proposed the strategy.

114. It appears that the (flamboyant magazine) editor who on Thursday evilly disparaged Stella reviewed the volume.

115. It appears that the (poor hog) farmer who on Friday openly disrespected Natalie culled the pigs.

116. It appears that the (radical German) filmmaker who on Saturday persistently distracted Emilia toured the set.

117. It appears that the (seasoned cod) fisherman who on Sunday wisely educated Addison released the swordfish.

118. It appears that the (bloodthirsty rhino) poacher who today dexterously evaded Lucy slayed the elephant.

119. It appears that the (learned private) optician who yesterday tenderly embraced Audrey examined the glasses.

120. It appears that the (pious Orthodox) rabbi who tonight highly endorsed Bella said the prayer.

121. It appears that the (traumatized ER) physician who last night kindly enlightened Brooklyn caught the disease.

122. It appears that the (courageous jet) pilot who this week courteously escorted Savannah flew the bomber.

123. It appears that the (swaggering navy) officer who last week undoubtedly exploited Claire navigated the boat.

124. It appears that the (aspiring state) politician who on Monday fully exposed Naomi announced the result.

125. It appears that the (controversial Libertarian) congresswoman who on Tuesday bravely faced Caroline headed the opposition.

126. It appears that the (unconventional Afghan) imam who on Wednesday tremendously flattered Carolina visited the mosque.

127. It appears that the (stressed brain) surgeon who on Thursday badly frightened Maya invented the cure.

128. It appears that the (careless Spanish) translator who on Friday bitterly denounced Valentina proofread the subtitles.

129. It appears that the (comforting Catholic) priest who on Saturday tightly hugged Ruby delivered the sermon.

130. It appears that the (sassy hotel) receptionist who on Sunday enduringly humored Kennedy rang the bell.

131. It appears that the (regretful former) president who today sexually harassed Ivy cleared the desk.

132. It appears that the (calculating insurance) CEO who yesterday profoundly humiliated Ariana filed the claim.

133. It appears that the (tireless warehouse) supervisor who tonight suddenly hit Allison rearranged the shelves.

134. It appears that the (prominent Japanese) geographer who last night perfectly imitated Gabriella predicted the earthquake.

135. It appears that the (competent plumbing) inspector who this week obediently indulged Samantha scrutinized the sink.

136. It appears that the (boring internal) auditor who last week unfairly assessed Hailey scanned the receipts.

137. It appears that the (affluent oil) investor who on Monday truly inspired Delilah funded the enterprise.

138. It appears that the (generous media) sponsor who on Tuesday brusquely interrupted Eva advertised the platform.

139. It appears that the (eccentric stamp) collector who on Wednesday hurriedly interviewed Piper tested the authenticity.

140. It appears that the (persistent sales) representative who on Thursday earnestly invited Sadie promoted the brand.

141. It appears that the (frugal wholesale) buyer who on Friday extremely irritated Josephine rented the van.

142. It appears that the (violent drug) dealer who on Saturday fondly kissed Adeline incited the fight.

143. It appears that the (sociable Dutch) merchant who on Sunday loudly lauded Emery exchanged the antiques.

144. It appears that the (gullible apartment) owner who today sternly lectured Lydia mailed the deposit.

145. It appears that the (persevering bread) baker who yesterday inappropriately touched Clara baked the sourdough.

146. It appears that the (extravagant wedding) florist who tonight speedily calmed Vivian trimmed the roses.

147. It appears that the (dubious magic) practitioner who last night consistently underestimated Peyton revealed the secret.

148. It appears that the (certified heart) specialist who this week formally nominated Julia diagnosed the illness.

149. It appears that the (muscular ambulance) paramedic who last week barely noticed Arya filled the syringe.

150. It appears that the (caring kindergarten) principal who on Monday duly notified Sansa vetted the syllabi.

151. It appears that the (uncompromising vice) dean who on Tuesday deeply offended Liam addressed the concerns.

152. It appears that the (fearless flight) instructor who on Wednesday strongly opposed Noah landed the helicopter.

153. It appears that the (brainy math) tutor who on Thursday completely outwitted Oliver explained the subtraction.

154. It appears that the (intelligent data) scientist who on Friday quietly passed William forecast the demand.

155. It appears that the (obsessed Shakespeare) scholar who on Saturday continually pestered Elijah perused the books.

156. It appears that the (valiant polar)explorer who on Sunday excitedly recruited James braved the cold.

157. It appears that the (flexible Chinese) gymnast who today shyly photographed Benjamin traversed the beam.

158. It appears that the (sleepless industrial) chemist who yesterday unhesitatingly picked Lucas combined the substances.

159. It appears that the (leading Renaissance) historian who tonight immensely pleased Henry authored the manuscript.

160. It appears that the (adventurous field) linguist who last night proudly presented Alexander learned the language.

161. It appears that the (bold Russian) astronaut who this week feebly protested Madison roamed the moon.

162. It appears that the (kooky zodiac) astrologer who last week valiantly protected Michael studied the stars.

163. It appears that the (brawny submarine) sailor who on Monday smugly provoked Ethan sank the battleship.

164. It appears that the (talkative news) correspondent who on Tuesday justly punished Jacob reported the incident.

165. It appears that the (industrious construction) worker who on Wednesday decisively fired Logan transported the bricks.

166. It appears that the (stoic stone) mason who on Thursday constantly terrified Jackson shaped the statue.

167. It appears that the (suspicious graveyard) caretaker who on Friday totally discredited Levi unearthed the coffin.

168. It appears that the (grumpy building) custodian who on Saturday instantly recognized Sebastian maintained the locks.

169. It appears that the (joyful party) host who on Sunday frantically phoned Mateo ordered the pizzas.

170. It appears that the (tall basketball) captain who today pitilessly ridiculed Jack sang the anthem.

171. It appears that the (feisty line) cook who yesterday miraculously rescued Owen grilled the fish.

172. It appears that the (infamous Somalian) pirate who tonight selflessly saved Yasir loaded the cannon.

173. It appears that the (welcoming travel) guide who last night evidently scared Aiden planned the itinerary.

174. It appears that the (fluent medical) interpreter who this week convincingly schooled Samuel forwarded the email.

175. It appears that the (unemployed coal) miner who last week furiously scolded Joseph uncaged the canary.

176. It appears that the (fatigued telephone) operator who on Monday cunningly selected David unplugged the cord.

177. It appears that the (daring fossil) hunter who on Tuesday foully slandered Wyatt excavated the site.

178. It appears that the (muddy Swedish) gardener who on Wednesday playfully slapped Carter planted the trees.

179. It appears that the (informative park) ranger who on Thursday treacherously stabbed Asher provided the guidelines.

180. It appears that the (vocal education) expert who on Friday cautiously stalked Julian advocated the policy.

181. It appears that the (overpaid management) consultant who on Saturday disagreeably startled Jayden prepared the slides.

182. It appears that the (rowdy Protestant) clergyman who on Sunday respectfully declined Ezra collected the tithe.

183. It appears that the (inspiring Presbyterian) pastor who today urgently summoned Joshua preached the gospel.

184. It appears that the (unsuccessful film) producer who yesterday closely supervised Nathan spearheaded the animation.

185. It appears that the (introverted Vietnamese) beekeeper who tonight jokingly tackled Caleb harvested the honey.

186. It appears that the (sarcastic book) reviewer who last night systematically trained Santiago penned the critiques.

187. It appears that the (determined sit-in) protester who this week visibly upset Ezekiel demanded the resignation.

188. It appears that the (speedy Olympic) sprinter who last week personally visited Isaiah accepted the medal.

189. It appears that the (attractive varsity) swimmer who on Monday timidly wooed Xavier failed the course.

190. It appears that the (tenacious marathon) runner who on Tuesday needlessly worried Kai bandaged the wound.

191. It appears that the (daredevil mountain) biker who on Wednesday brilliantly surpassed Jose lost the race.

192. It appears that the (entertaining variety) streamer who on Thursday actively engaged Silas streamed the show.

193. It appears that the (stubborn AIDS) denier who on Friday mischievously misled Amir distrusted the government.

194. It appears that the (motivated contract) lobbyist who on Saturday cruelly deceived Enzo cast the vote.

195. It appears that the (enthusiastic day) hiker who on Sunday craftily fooled Omar unfolded the map.

196. It appears that the (arrogant urban) planner who today shockingly cheated Javier hated the downtown.

197. It appears that the (cautious narcotics) detective who yesterday immorally overcharged Griffin infiltrated the gang.

198. It appears that the (impoverished Syrian) refugee who tonight easily seduced Zain boarded the boat.

199. It appears that the (courteous parking) valet who last night remoreselessly chastised Ramon fetched the limousine.

200. It appears that the (dusty street) cleaner who this week certainly enraged Remi swept the boulevard.

APPENDIX E

STIMULI SENTENCES FOR EXPERIMENT 4A/4B

Sentences from Experiment 4a do not contain the *it*-cleft (*It was*) and the complementizer *that*. Sentences from Experiment 4b contain both the underlined materials. Experimental items with similar and dissimilar coordinated referents are presented separately in E1 and E2, respectively. All experimental items presented below have three coordinated referents. To form conditions with two coordinated referents, remove the determiner and the first noun in parentheses (*the hamster* in sentence 1 in E1). To form conditions with only one referent, remove all materials in parentheses, both nouns, their determiners and *and* (*the hamster, the cat and* in sentence 1 in E1).

E.1 Experimental Items (Similar Referents)

1. It was (the hamster, the cat and) the dog that, according to Mary, had been recently abandoned by the tenant.

2. It was (the lion, the leopard and) the tiger that, according to John, had been frequently purchased by the collector.

3. It was (the eagle, the hawk and) the falcon that, according to Judith, had been repeatedly spotted by the birdwatcher.

4. It was (the shark, the dolphin and) the whale that, according to Adam, had been carefully drawn by the kindergartener.

5. It was (the cow, the pig and) the chicken that, according to Jin, had been recently bought by the millionaire.

6. It was (the swan, the goose and) the duck that, according to Jane, had been carefully sketched by the student.

7. It was (the mule, the pony and) the donkey that, according to Simon, had been recently sold by the auctioneer.

8. It was (the mosquito, the bee and) the ant that, according to Elena, had been meticulously painted by the artist.

9. It was (the robin, the dove and) the pigeon that, according to Ariana, had been repeatedly filmed by the director.

10. It was (the owl, the vulture and) the crow that, according to Grant, had been adequately described by the poet.

11. It was (the crab, the lobster and) the shrimp that, according to Anne, had been meticulously studied by the zoologist.

12. It was (the snake, the frog and) the toad that, according to Juliette, had been constantly weighed by the owner.

13. It was (the chair, the table and) the desk that, according to William, had been carelessly measured by the carpenter.

14. It was (the ceiling, the door and) the window that, according to Abby, had been repeatedly photographed by the photographer.

15. It was (the blanket, the pillow and) the mattress that, according to Katie, had been severely destroyed by the fire.

16. It was (the chandelier, the candle and) the lamp that, according to Liz, had been recently found by the neighbor.

17. It was (the drawer, the wardrobe and) the armoire that, according to Bruce, had been severely hit by the bullets.

18. It was (the cupboard, the microwave and) the oven that, according to Claire, had been sufficiently inspected by the chef.

19. It was (the computer, the laptop and) the phone that, according to Jeff, had been recently smashed by the creature.

20. It was (the painting, the poster and) the map that, according to Isaiah, had been constantly repositioned by the designer.

21. It was (the DVD, the magazine and) the book that, according to Lauren, had been constantly discussed by the organizer.

22. It was (the jug, the jar and) the vase that, according to Jake, had been carelessly unearthed by the archeologist.

23. It was (the steamer, the pot and) the pan that, according to Simone, had been adequately examined by the cook.

24. It was (the bathtub, the shower and) the sink that, according to Chris, had been meticulously scrutinized by the investigator.

25. It was (the hallway, the entrance and) the exit that, according to David, had been intermittently monitored by the security guard.

26. It was (the gallery, the museum and) the bookstore that, according to Ella, had been severely ravaged by the earthquake.

27. It was (the garden, the zoo and) the park that, according to Angela, had been accidentally discovered by the tourist.

28. It was (the bedroom, the kitchen and) the lounge that, according to Zoe, had been thoroughly searched by the police.

29. It was (the library, the monument and) the fountain that, according to Cate, had been loosely depicted by the novelist.

30. It was (the armory, the castle and) the palace that, according to Nicole, had been repeatedly damaged by the siege.

31. It was (the house, the condominium and) the skyscraper that, according to John, had been severely burned by the arsonist.

32. It was (the market, the diner and) the restaurant that, according to Maddie, had been adequately disinfected by the technician.

33. It was (the theater, the bar and) the club that, according to Erik, had been repeatedly reserved by the CEO.

34. It was (the store, the supermarket and) the mall that, according to Jackson, had been sloppily designed by the architect.

35. It was (the hospital, the pharmacy and) the clinic that, according to Alex, had been forcefully acquired by the city.

36. It was (the synagogue, the mosque and) the church that, according to Tobias, had been constantly seized by the rebels.

E.2 Experimental Items (Dissimilar Referents)

1. It was (the car, the TV and) the dog that, according to Mary, had been recently abandoned by the tenant.

2. It was (the sofa, the bicycle and) the tiger that, according to John, had been frequently purchased by the collector.

3. It was (the star, the plane and) the falcon that, according to Judith, had been repeatedly spotted by the birdwatcher.

4. It was (the bed, the sky and) the whale that, according to Adam, had been carefully drawn by the kindergartener.

5. It was (the bus, the futon and) the chicken that, according to Jin, had been recently bought by the millionaire.

6. It was (the reed, the river and) the duck that, according to Jane, had been carefully sketched by the student.

7. It was (the tractor, the pond and) the donkey that, according to Simon, had been recently sold by the auctioneer.

8. It was (the sunflower, the garage and) the ant that, according to Elena, had been meticulously painted by the artist.

9. It was (the bathroom, the concert and) the pigeon that, according to Ariana, had been repeatedly filmed by the director.

10. It was (the glacier, the boat and) the crow that, according to Grant, had been adequately described by the poet.

11. It was (the sand, the seaweed and) the shrimp that, according to Anne, had been meticulously studied by the zoologist.

12. It was (the aquarium, the mushroom and) the toad that, according to Juliette, had been constantly weighed by the owner.

13. It was (the oak, the yard and) the desk that, according to William, had been carelessly measured by the carpenter.

14. It was (the scooter, the elm and) the window that, according to Abby, had been repeatedly photographed by the photographer.

15. It was (the fir, the attic and) the mattress that, according to Katie, had been severely destroyed by the fire.

16. It was (the lizard, the banana and) the lamp that, according to Liz, had been recently found by the neighbor.

17. It was (the skateboard, the rabbit and) the armoire that, according to Bruce, had been severely hit by the bullets.

18. It was (the dining room, the fish and) the oven that, according to Claire, had been sufficiently inspected by the chef.

19. It was (the office, the orchid and) the phone that, according to Jeff, had been recently smashed by the creature.

20. It was (the turtle, the basil and) the map that, according to Isaiah, had been constantly repositioned by the designer.

21. It was (the conference, the venue and) the book that, according to Lauren, had been constantly discussed by the organizer.

22. It was (the flower, the skeleton and) the vase that, according to Jake, had been carelessly unearthed by the archeologist.

23. It was (the storeroom, the meat and) the pan that, according to Simone, had been adequately examined by the cook.

24. It was (the blood, the patio and) the sink that, according to Chris, had been meticulously scrutinized by the investigator.

25. It was (the van, the money and) the exit that, according to David, had been intermittently monitored by the security guard.

26. It was (the train, the willow and) the bookstore that, according to Ella, had been severely ravaged by the earthquake.

27. It was (the elephant, the pine and) the park that, according to Angela, had been accidentally discovered by the tourist.

28. It was (the SUV, the fridge and) the lounge that, according to Zoe, had been thoroughly searched by the police.

29. It was (the raccoon, the ambulance and) the fountain that, according to Cate, had been loosely depicted by the novelist.

30. It was (the tank, the maple and) the palace that, according to Nicole, had been repeatedly damaged by the siege.

31. It was (the moped, the bush and) the skyscraper that, according to John, had been severely burned by the arsonist.

32. It was (the truck, the box and) the restaurant that, according to Maddie, had been adequately disinfected by the technician.

33. It was (the helicopter, the projector and) the club that, according to Erik, had been repeatedly reserved by the CEO.

34. It was (the ship, the bench and) the mall that, according to Jackson, had been sloppily designed by the architect.

35. It was (the airplane, the beech and) the clinic that, according to Alex, had been forcefully acquired by the city.

36. It was (the food, the cannon and) the church that, according to Tobias, had been constantly seized by the rebels.

E.3 Filler items

1. It was the buffalo that, according to Mary, was grazing the field next to the farmer.

2. It was the engineer that, according to Grant, was fixing the pipe below the machine.

3. It was the musician that, according to Anne, was practicing the verse with the singer.

4. It was the actor that, according to Juliette, was rehearsing the play with the director.

5. It was the choir that, according to John, was singing the hymn in the chapel.

6. It was the teenager that, according to Judith, was riding the roller coaster at the fair.

7. It was the eruption that, according to Adam, was destroying the villages in the south.

8. It was the heat that, according to Jin, was killing the bacteria in the steak.

9. It was the flood that, according to Jane, was drowning the cities in the north.

10. It was the rain that, according to Simon, was soaking the clothes on the line.
11. It was the noise that, according to Elena, was vibrating the glass of the building.
12. It was the scientist that, according to Ariana, was reading the paper on the couch.
13. It was the jaguar and the cougar that, according to William, were chewing the bones next to the zoologist.
14. It was the vendor and the customer that, according to Jake, were appraising the gem in the store.
15. It was the lawyer and the sheriff that, according to Simone, were accusing the suspect of the crime.
16. It was the politician and the secretary that, according to Chris, were proposing the policy to the senator.
17. It was the fire and the smoke that, according to Abby, were engulfing the towns around the capital.
18. It was the tsunami and the hurricane that, according to Katie, were disrupting the economy of the country.
19. It was the finch and the woodpecker that, according to Liz, were drinking the water from the pond.
20. It was the octopus and the squid that, according to Bruce, were chasing the prey through the coral.
21. It was the ferry and the canoe that, according to Claire, were following the seagulls in the sky.
22. It was the rat and the mouse that, according to Jeff, were gnawing the cheese in the sewer.

23. It was the fox and the wolf that, according to Isaiah, were fleeing the hunter with the rifle.

24. It was the soldier and the pilot that, according to Lauren, were defending the fort near the border.

25. It was the donor, the magnate and the intermediary that, according to David, were debating the issue at the party.

26. It was the headmistress, the teacher and the parent that, according to Jackson, were encouraging the student at the meeting.

27. It was the father, the mother and the child that, according to Alex, were cooking the dish from the cookbook.

28. It was the sloth, the armadillo and the platypus that, according to Tobias, were sharing the treats from the zookeeper.

29. It was the conservationist, the poacher and the ranger that, according to Ella, were signing the agreement at the conference.

30. It was the gecko, the iguana and the tortoise that, according to Angela, were enjoying the sun in the summer.

31. It was the accountant, the waiter and the manager that, according to Zoe, were browsing the products in the shop.

32. It was the bear, the moose and the elk that, according to Cate, were crossing the bridge over the creek.

33. It was the antelope, the goat and the sheep that, according to Nicole, were climbing the hill near the pasture.

34. It was the hare, the deer and the squirrel that, according to John, were rolling the acorns down the hill.

35. It was the carpenter, the potter and the electrician that, according to Maddie, were paying the barber for the service.

36. It was the chihuahua, the dalmatian and the poodle that, according to Erik, were fetching the toy from the woman.

APPENDIX F
STIMULI SENTENCES FOR EXPERIMENT 5

Experimental items with similar and dissimilar coordinated referents are presented separately in F1 and F2, respectively. All experimental items presented below have three coordinated referents. To form conditions with two coordinated referents, remove the determiner and the first noun in parentheses (*the hamster,* in sentence 1 in F1).

F.1 Experimental Items (Similar Referents)

1. It was (the hamster,) the cat and the dog that, according to Andy, were abandoned by the tenant.

2. It was (the lion,) the leopard and the tiger that, according to Amy, were purchased by the collector.

3. It was (the eagle,) the hawk and the falcon that, according to Ben, were spotted by the birdwatcher.

4. It was (the shark,) the dolphin and the whale that, according to Luke, were drawn by the kindergartener.

5. It was (the cow,) the pig and the rooster that, according to Kevin, were bought by the millionaire.

6. It was (the swan,) the goose and the duck that, according to Leah, were sketched by the student.

7. It was (the mule,) the pony and the donkey that, according to Jack, were sold by the auctioneer.

8. It was (the mosquito,) the bee and the ant that, according to Sophia, were painted by the artist.

9. It was (the mallard,) the dove and the pigeon that, according to Matthew, were filmed by the director.
10. It was (the owl,) the vulture and the crow that, according to Tom, were described by the poet.
11. It was (the crab,) the lobster and the shrimp that, according to Stephanie, were studied by the zoologist.
12. It was (the tortoise,) the frog and the toad that, according to Becky, were weighed by the owner.
13. It was (the chair,) the table and the desk that, according to Ellen, were measured by the carpenter.
14. It was (the ceiling,) the door and the window that, according to Anna, were photographed by the photographer.
15. It was (the blanket,) the pillow and the mattress that, according to Teddy, were destroyed by the hurricane.
16. It was (the chandelier,) the candle and the lamp that, according to Alex, were found by the neighbor.
17. It was (the drawer,) the wardrobe and the armoire that, according to Zack, were hit by the bullets.
18. It was (the stove,) the microwave and the oven that, according to Madeline, were inspected by the chef.
19. It was (the computer,) the laptop and the phone that, according to Sasha, were smashed by the creature.
20. It was (the painting,) the poster and the picture that, according to Fiona, were repositioned by the designer.

21. It was (the DVD,) the CD and the Blu-ray that, according to Ivan, were discussed by the organizer.

22. It was (the vase,) the jar and the jug that, according to Issac, were unearthed by the archeologist.

23. It was (the wok,) the pot and the pan that, according to Alma, were examined by the cook.

24. It was (the bathtub,) the shower and the sink that, according to Ian, were scrutinized by the investigator.

25. It was (the hallway,) the entrance and the staircase that, according to Rosie, were monitored by the security guard.

26. It was (the bookstore,) the museum and the library that, according to Isabella, were ravaged by the flood.

27. It was (the garden,) the zoo and the park that, according to Jessica, were discovered by the tourist.

28. It was (the bedroom,) the kitchen and the lounge that, according to Tess, were searched by the police.

29. It was (the statue,) the monument and the fountain that, according to Keith, were depicted by the novelist.

30. It was (the citadel,) the castle and the palace that, according to Simon, were damaged by the siege.

31. It was (the apartment,) the condominium and the townhouse that, according to Elena, were burned by the arsonist.

32. It was (the steakhouse,) the diner and the restaurant that, according to Aliza, were disinfected by the technician.

33. It was (the ballroom,) the cabaret and the nightclub that, according to Justin, were reserved by the CEO.

34. It was (the store,) the supermarket and the mall that, according to Daniel, were designed by the architect.

35. It was (the hospital,) the pharmacy and the clinic that, according to Fiora, were acquired by the county.

36. It was (the synagogue,) the mosque and the church that, according to Lin, were seized by the rebels.

37. It was (the sword,) the lance and the shield that, according to Ross, were melted by the blacksmith.

38. It was (the sandwich,) the burger and the fries that, according to Rachel, were poisoned by the psychopath.

39. It was (the sun,) the moon and the sky that, according to Monica, were observed by the scientist.

40. It was (the egg,) the cheese and the milk that, according to Chandler, were exported by the company.

41. It was (the spinach,) the kale and the lettuce that, according to Joey, were imported by the region.

42. It was (the turnip,) the carrot and the potato that, according to Phoebe, were washed by the worker.

43. It was (the onion,) the ginger and the garlic that, according to Shen, were checked by the grocer.

44. It was (the apple,) the pear and the plum that, according to Janice, were cut by the maid.

45. It was (the lemon,) the lime and the orange that, according to John, were delivered by the mailman.

46. It was (the strawberry,) the blueberry and the cherry that, according to Lenore, were swallowed by the child.

47. It was (the taco,) the burrito and the quesadilla that, according to Jason, were made by the restauranteur.

48. It was (the salad,) the soup and the dessert that, according to Laura, were prepared by the landlady.

49. It was (the cake,) the pie and the tart that, according to Rodrigo, were devoured by the stranger.

50. It was (the quiche,) the pancake and the waffle that, according to Eszter, were sent by the grandmother.

51. It was (the broccoli,) the cauliflower and the cabbage that, according to Mike, were roasted by the servant.

52. It was (the wasp,) the spider and the cockroach that, according to Aurora, were watched by the kid.

53. It was (the heron,) the stork and the crane that, according to Han, were shot by the hunter.

54. It was (the anaconda,) the python and the snake that, according to Ilya, were confiscated by the cop.

55. It was (the chameleon,) the gecko and the iguana that, according to Rich, were hidden by the thief.

56. It was (the rat,) the mouse and the hedgehog that, according to Grace, were collected by the king.

57. It was (the kangaroo,) the koala and the platypus that, according to Emre, were smuggled by the poacher.

58. It was (the elephant,) the zebra and the giraffe that, according to Joan, were sought by the explorer.

59. It was (the poodle,) the dalmatian and the beagle that, according to Jaren, were pursued by the soldier.

60. It was (the gorilla,) the chimp and the monkey that, according to Harold, were neglected by the zookeeper.

61. It was (the beaver,) the chipmunk and the squirrel that, according to Harry, were ignored by the vet.

62. It was (the panther,) the jaguar and the cougar that, according to Emma, were buried by the earthquake.

63. It was (the elk,) the caribou and the sheep that, according to Fred, were unveiled by the magician.

64. It was (the clam,) the scallop and the mussel that, according to George, were trapped by the net.

65. It was (the squid,) the octopus and the oyster that, according to Mia, were dissected by the biologist.

66. It was (the snail,) the slug and the jellyfish that, according to Jules, were feared by the surfer.

67. It was (the knife,) the fork and the spoon that, according to Gaby, were dropped by the highschooler.

68. It was (the glass,) the mug and the cup that, according to Omer, were stained by the dye.

69. It was (the T-shirt,) the cardigan and the polo that, according to Carlos, were soaked by the downpour.

70. It was (the pizza,) the spaghetti and the lasagna that, according to Caitlyn, were discarded by the son.

71. It was (the rice,) the noodles and the bread that, according to Garen, were donated by the tycoon.

72. It was (the sundae,) the gelato and the sorbet that, according to Hans, were lauded by the critic.

73. It was (the muffin,) the donut and the cupcake that, according to Jonie, were decorated by the pastor.

74. It was (the beef,) the pork and the chicken that, according to Paul, were accepted by the beggar.

75. It was (the butter,) the oil and the margarine that, according to Minnie, were stored by the fiance.

76. It was (the mayonnaise,) the ketchup and the mustard that, according to Kate, were hoarded by the teenager.

77. It was (the kiwi,) the mango and the pineapple that, according to Charles, were taken by the mugger.

78. It was (the bacon,) the sausage and the ham that, according to Kwame, were stolen by the cousin.

79. It was (the steak,) the ribs and the brisket that, according to Lydia, were supplied by the organization.

80. It was (the cookie,) the brownie and the croissant that, according to Albert, were obtained by the refugee.

81. It was (the peach,) the raspberry and the melon that, according to Nelson, were received by the patient.

82. It was (the sushi,) the ramen and the pho that, according to Andrew, were praised by the reviewer.

83. It was (the thermos,) the bottle and the flask that, according to Mark, were given by the donor.

84. It was (the hat,) the cap and the beanie that, according to Vincent, were contributed by the trustee.

85. It was (the dress,) the blouse and the skirt that, according to Flora, were presented by the salesperson.

86. It was (the tie,) the scarf and the belt that, according to Derek, were gifted by the uncle.

87. It was (the boots,) the sandals and the heels that, according to Olivia, were provided by the boss.

88. It was (the coat,) the sweater and the jacket that, according to Ava, were preferred by the retiree.

89. It was (the jeans,) the shorts and the pants that, according to Mia, were chosen by the committee.

90. It was (the shoes,) the socks and the stockings that, according to Charlotte, were selected by the orphan.

91. It was (the purse,) the wallet and the bag that, according to Amelia, were manufactured by the state.

92. It was (the shampoo,) the conditioner and the soap that, according to Harper, were crafted by the artisan.

93. It was (the comb,) the scrub and the brush that, according to Evelyn, were broken by the explosion.

94. It was (the plate,) the bowl and the saucer that, according to Abigail, were thrown by the maniac.

95. It was (the tea,) the coffee and the chocolate that, according to Alexis, were licked by the infant.

96. It was (the wine,) the liquor and the beer that, according to Emily, were touched by the toddler.

97. It was (the whisky,) the rum and the vodka that, according to Ella, were smelled by the father.

98. It was (the cloud,) the rain and the fog that, according to Camille, were predicted by the forecaster.

99. It was (the stone,) the pebble and the boulder that, according to Aria, were hurled by the pitcher.

100. It was (the bronze,) the silver and the gold that, according to Penelope, were targeted by the pirate.

101. It was (the emerald,) the pearl and the jade that, according to Avery, were retrieved by the family.

102. It was (the mountain,) the hill and the valley that, according to Mila, were loved by the people.

103. It was (the lake,) the sea and the creek that, according to Scarlett, were liked by the locals.

104. It was (the guitar,) the cello and the violin that, according to Layla, were transported by the driver.

105. It was (the trumpet,) the trombone and the saxophone that, according to Chloe, were investigated by the detective.

106. It was (the flute,) the clarinet and the bassoon that, according to Victoria, were hauled by the musician.

107. It was (the diamond,) the ruby and the sapphire that, according to Eleanor, were returned by the criminal.

108. It was (the iron,) the aluminum and the zinc that, according to Nora, were noticed by the pedestrian.

109. It was (the avocado,) the papaya and the grape that, according to Lucy, were bagged by the cashier.

110. It was (the apricot,) the prune and the fig that, according to Audrey, were procured by the intermediary.

111. It was (the cucumber,) the eggplant and the zucchini that, according to Bella, were sliced by the aunt.

112. It was (the tangerine,) the mandarin and the watermelon that, according to Brooklyn, were carried by the shopper.

113. It was (the axe,) the hatchet and the chainsaw that, according to Savannah, were showcased by the lumberjack.

114. It was (the screw,) the wrench and the drill that, according to Claire, were brought by the friend.

115. It was (the ink,) the pen and the sharpie that, according to Naomi, were held by the principal.

116. It was (the needle,) the yarn and the thread that, according to Caroline, were produced by the seamstress.

117. It was (the ottoman,) the couch and the stool that, according to Gabriella, were moved by the renter.

118. It was (the hammock,) the cot and the bunk that, according to Samantha, were assembled by the daughter.

119. It was (the bookcase,) the closet and the rack that, according to Hailey, were cursed by the witch.

120. It was (the chimney,) the roof and the balcony that, according to Delilah, were cleaned by the butler.

121. It was (the smoke,) the ash and the fire that, according to Eva, were detected by the firefighter.

122. It was (the farm,) the barn and the pasture that, according to Piper, were tidied by the farmhand.

123. It was (the earring,) the necklace and the brooch that, according to Sadie, were located by the husband.

124. It was (the stamp,) the envelope and the letter that, according to Josephine, were misplaced by the acquaintance.

125. It was (the tire,) the headlight and the bumper that, according to Sansa, were fixed by the mechanic.

126. It was (the leaf,) the stem and the vine that, according to Liam, were bent by the beast.

127. It was (the lotion,) the perfume and the deodorant that, according to Noah, were used by the influencer.

128. It was (the lipstick,) the mascara and the eyeliner that, according to Oliver, were promoted by the model.

129. It was (the curtain,) the drape and the sash that, according to Elijah, were hung by the janitor.

130. It was (the toothbrush,) the mouthwash and the floss that, according to William, were recommended by the dentist.

131. It was (the coil,) the cable and the wire that, according to James, were repaired by the electrician.

132. It was (the cord,) the socket and the plug that, according to Benjamin, were kept by the handyman.

133. It was (the towel,) the washcloth and the napkin that, according to Lucas, were ordered by the nephew.

134. It was (the avenue,) the street and the road that, according to Levi, were scoured by the bandits.

135. It was (the tuna,) the salmon and the cod that, according to Sebastian, were speared by the fisherman.

136. It was (the boar,) the moose and the deer that, according to Mateo, were worshipped by the cavepeople.

137. It was (the nutmeg,) the caramel and the vanilla that, according to Jack, were relished by the expert.

138. It was (the rhino,) the buffalo and the bison that, according to Owen, were attacked by the smugglers.

139. It was (the coyote,) the fox and the wolf that, according to Theodore, were detested by the rancher.

140. It was (the driveway,) the courtyard and the porch that, according to Aiden, were rearranged by the gardener.

141. It was (the lantern,) the flashlight and the torch that, according to Samuel, were forgotten by the guest.

142. It was (the harbor,) the pier and the lighthouse that, according to Joseph, were safeguarded by the patrol.

143. It was (the missile,) the grenade and the bomb that, according to Nathan, were solicited by the insurrectionists.

144. It was (the helmet,) the goggles and the mask that, according to Caleb, were demanded by the authority.

145. It was (the wool,) the feather and the fur that, according to Santiago, were stockpiled by the aristocrat.

146. It was (the linen,) the fabric and the cloth that, according to Ezekiel, were fetched by the chauffeur.

147. It was (the bolt,) the lock and the knob that, according to Brooks, were rejected by the locksmith.

148. It was (the tile,) the concrete and the brick that, according to Xavier, were evaluated by the builder.

149. It was (the pea,) the bean and the pumpkin that, according to Kai, were fancied by the lawyer.

150. It was (the dart,) the frisbee and the ball that, according to Jose, were grabbed by the referee.

151. It was (the movie,) the play and the show that, according to Riley, were appreciated by the playwright.

152. It was (the meth,) the heroin and the cocaine that, according to Zoey, were contained by the specialist.

153. It was (the virus,) the fungus and the bacteria that, according to Hannah, were avoided by the nurse.

154. It was (the breakfast,) the lunch and the dinner that, according to Hazel, were shared by the household.

155. It was (the novel,) the poem and the story that, according to Violet, were complimented by the professor.

156. It was (the tooth,) the hair and the nail that, according to Stella, were extracted by the paleontologist.

157. It was (the ice,) the snow and the water that, according to Natalie, were removed by the cleaner.

158. It was (the pudding,) the jello and the panna cotta that, according to Emilia, were left by the attendee.

159. It was (the jelly,) the jam and the marmalade that, according to Addison, were boxed by the seller.

160. It was (the pitchfork,) the shovel and the rake that, according to Carolina, were rated by the farmer.

161. It was (the shotgun,) the revolver and the pistol that, according to Maya, were swung by the enemy.

162. It was (the wood,) the plastic and the metal that, according to Valentina, were gathered by the vagabond.

163. It was (the leather,) the velvet and the silk that, according to Ruby, were replaced by the producer.

164. It was (the gauze,) the tweezers and the bandage that, according to Kennedy, were distributed by the government.

165. It was (the brook,) the stream and the waterfall that, according to Ivy, were guarded by the ranger.

166. It was (the village,) the city and the town that, according to Ariana, were protected by the military.

167. It was (the cliff,) the island and the peninsula that, according to Allison, were submerged by the tsunami.

168. It was (the keyboard,) the joystick and the screen that, according to Adeline, were exchanged by the gamer.

169. It was (the camera,) the lens and the tripod that, according to Emery, were auctioned by the celebrity.

170. It was (the highlighter,) the sharpener and the pencil that, according to Lydia, were advertised by the firm.

171. It was (the crayon,) the paint and the easel that, according to Clara, were exhibited by the painter.

172. It was (the compass,) the binoculars and the telescope that, according to Vivian, were displayed by the librarian.

173. It was (the rug,) the carpet and the flooring that, according to Peyton, were dried by the heat.

174. It was (the tray,) the basket and the bucket that, according to Julia, were borrowed by the relative.

175. It was (the game,) the trivia and the puzzle that, according to Arya, were enjoyed by the enthusiast.

176. It was (the corn,) the barley and the wheat that, according to Henry, were ground by the villager.

177. It was (the pepper,) the sugar and the salt that, according to Alexander, were separated by the peasant.

178. It was (the lemonade,) the juice and the soda that, according to Casey, were dispensed by the warden.

179. It was (the cinnamon,) the syrup and the honey that, according to Michael, were craved by the runner.

180. It was (the almond,) the cashew and the peanut that, according to Ethan, were crushed by the manufacturer.

181. It was (the magazine,) the tabloid and the newspaper that, according to Jacob, were surrendered by the niece.

182. It was (the dome,) the fortress and the tower that, according to Logan, were defended by the troops.

183. It was (the crossing,) the tunnel and the bridge that, according to Jackson, were surrounded by the barriers.

184. It was (the deli,) the bakery and the cafe that, according to David, were cherished by the traveler.

185. It was (the pub,) the tavern and the brewery that, according to Wyatt, were hated by the community.

186. It was (the hotel,) the motel and the inn that, according to Carter, were bombed by the cultists.

187. It was (the cabin,) the hut and the tent that, according to Asher, were disliked by the backpackers.

188. It was (the bamboo,) the coconut and the palm that, according to Julian, were relocated by the planner.

189. It was (the beet,) the celery and the asparagus that, according to Jayden, were needed by the host.

190. It was (the microphone,) the earbuds and the headphones that, according to Ezra, were desired by the girl.

191. It was (the guacamole,) the queso and the salsa that, according to Joshua, were requested by the boy.

192. It was (the eraser,) the blackboard and the chalk that, according to Silas, were snatched by the teacher.

193. It was (the manuscript,) the diary and the journal that, according to Amir, were owned by the prisoner.

194. It was (the malt,) the lager and the ale that, according to Enzo, were packed by the porter.

195. It was (the bourbon,) the brandy and the cider that, according to Omar, were wasted by the employer.

196. It was (the shuttle,) the streetcar and the subway that, according to Javier, were introduced by the governor.

197. It was (the coast,) the ocean and the shore that, according to Griffin, were misused by the investor.

198. It was (the lavender,) the clover and the poppy that, according to Zain, were treasured by the researcher.

199. It was (the jungle,) the woods and the forest that, according to Ramon, were harmed by the war.

200. It was (the hummus,) the pita and the naan that, according to Remi, were offered by the vendor.

F.2 Experimental Items (Dissimilar Referents)

1. It was (the kayak,) the TV and the dog that, according to Andy, were abandoned by the tenant.

2. It was (the sofa,) the bicycle and the tiger that, according to Amy, were purchased by the collector.

3. It was (the comet,) the plane and the falcon that, according to Ben, were spotted by the birdwatcher.

4. It was (the bed,) the star and the whale that, according to Luke, were drawn by the kindergartener.

5. It was (the bus,) the futon and the rooster that, according to Kevin, were bought by the millionaire.

6. It was (the reed,) the mound and the duck that, according to Leah, were sketched by the student.

7. It was (the car,) the pond and the donkey that, according to Jack, were sold by the auctioneer.

8. It was (the sunflower,) the garage and the ant that, according to Sophia, were painted by the artist.

9. It was (the bathroom,) the concert and the pigeon that, according to Matthew, were filmed by the director.

10. It was (the glacier,) the boat and the crow that, according to Tom, were described by the poet.

11. It was (the sand,) the elm and the shrimp that, according to Stephanie, were studied by the zoologist.

12. It was (the sandbox,) the banana and the toad that, according to Becky, were weighed by the owner.

13. It was (the oak,) the yard and the desk that, according to Ellen, were measured by the carpenter.

14. It was (the scooter,) the seaweed and the window that, according to Anna, were photographed by the photographer.

15. It was (the pole,) the attic and the mattress that, according to Teddy, were destroyed by the hurricane.

16. It was (the lizard,) the mushroom and the lamp that, according to Alex, were found by the neighbor.

17. It was (the skateboard,) the rabbit and the armoire that, according to Zack, were hit by the bullets.

18. It was (the storeroom,) the fish and the oven that, according to Madeline, were inspected by the chef.

19. It was (the factory,) the orchid and the phone that, according to Sasha, were smashed by the creature.

20. It was (the turtle,) the basil and the picture that, according to Fiona, were repositioned by the designer.

21. It was (the conference,) the billionaire and the Blu-ray that, according to Ivan, were discussed by the organizer.

22. It was (the flower,) the skeleton and the jug that, according to Issac, were unearthed by the archeologist.

23. It was (the backroom,) the herbs and the pan that, according to Alma, were examined by the cook.

24. It was (the blood,) the stable and the sink that, according to Ian, were scrutinized by the investigator.

25. It was (the van,) the money and the staircase that, according to Rosie, were monitored by the security guard.

26. It was (the train,) the willow and the library that, according to Isabella, were ravaged by the flood.

27. It was (the armadillo,) the pine and the park that, according to Jessica, were discovered by the tourist.

28. It was (the SUV,) the freezer and the lounge that, according to Tess, were searched by the police.

29. It was (the raccoon,) the ambulance and the fountain that, according to Keith, were depicted by the novelist.

30. It was (the tank,) the maple and the palace that, according to Simon, were damaged by the siege.

31. It was (the moped,) the bush and the townhouse that, according to Elena, were burned by the arsonist.

32. It was (the truck,) the box and the restaurant that, according to Aliza, were disinfected by the technician.

33. It was (the helicopter,) the projector and the nightclub that, according to Justin, were reserved by the CEO.

34. It was (the ship,) the bench and the mall that, according to Daniel, were designed by the architect.

35. It was (the airplane,) the tree and the clinic that, according to Fiora, were acquired by the county.

36. It was (the food,) the cannon and the church that, according to Lin, were seized by the rebels.

37. It was (the crowbar,) the popsicle and the shield that, according to Ross, were melted by the blacksmith.

38. It was (the air,) the vat and the fries that, according to Rachel, were poisoned by the psychopath.

39. It was (the llama,) the beech and the sky that, according to Monica, were observed by the scientist.

40. It was (the hammer,) the catfish and the milk that, according to Chandler, were exported by the company.

41. It was (the screwdriver,) the e-bike and the lettuce that, according to Joey, were imported by the region.

42. It was (the apron,) the counter and the potato that, according to Phoebe, were washed by the worker.

43. It was (the container,) the AC and the garlic that, according to Shen, were checked by the grocer.

44. It was (the packaging,) the carp and the plum that, according to Janice, were cut by the maid.

45. It was (the halibut,) the scissors and the orange that, according to John, were delivered by the mailman.

46. It was (the clip,) the insect and the cherry that, according to Lenore, were swallowed by the child.

47. It was (the souvenir,) the bust and the quesadilla that, according to Jason, were made by the restauranteur.

48. It was (the house,) the document and the dessert that, according to Laura, were prepared by the landlady.

49. It was (the algae,) the quail and the tart that, according to Rodrigo, were devoured by the stranger.

50. It was (the daisy,) the undershirt and the waffle that, according to Eszter, were sent by the grandmother.

51. It was (the bird,) the powder and the cabbage that, according to Mike, were roasted by the servant.

52. It was (the kite,) the rainbow and the cockroach that, according to Aurora, were watched by the kid.

53. It was (the conservationist,) the jeep and the crane that, according to Han, were shot by the hunter.

54. It was (the gun,) the glove and the snake that, according to Ilya, were confiscated by the cop.

55. It was (the rifle,) the woman and the iguana that, according to Rich, were hidden by the thief.

56. It was (the coin,) the truffle and the hedgehog that, according to Grace, were collected by the king.

57. It was (the drug,) the tractor and the platypus that, according to Emre, were smuggled by the poacher.

58. It was (the treasure,) the chariot and the giraffe that, according to Joan, were sought by the explorer.

59. It was (the robber,) the cash and the beagle that, according to Jaren, were pursued by the soldier.

60. It was (the patron,) the leakage and the monkey that, according to Harold, were neglected by the zookeeper.

61. It was (the disease,) the scream and the squirrel that, according to Harry, were ignored by the vet.

62. It was (the center,) the stump and the cougar that, according to Emma, were buried by the earthquake.

63. It was (the assistant,) the wand and the sheep that, according to Fred, were unveiled by the magician.

64. It was (the gem,) the diver and the mussel that, according to George, were trapped by the net.

65. It was (the heart,) the twig and the oyster that, according to Mia, were dissected by the biologist.

66. It was (the pollution,) the freighter and the jellyfish that, according to Jules, were feared by the surfer.

67. It was (the pet,) the plant and the spoon that, according to Gaby, were dropped by the highschooler.

68. It was (the floor,) the shirt and the cup that, according to Omer, were stained by the dye.

69. It was (the possum,) the secretary and the polo that, according to Carlos, were soaked by the downpour.

70. It was (the card,) the sneakers and the lasagna that, according to Caitlyn, were discarded by the son.

71. It was (the quilt,) the shelter and the bread that, according to Garen, were donated by the tycoon.

72. It was (the decor,) the waitress and the sorbet that, according to Hans, were lauded by the critic.

73. It was (the altar,) the gown and the cupcake that, according to Jonie, were decorated by the pastor.

74. It was (the check,) the hoodie and the chicken that, according to Paul, were accepted by the beggar.

75. It was (the chest,) the cart and the margarine that, according to Minnie, were stored by the fiance.

76. It was (the toothpaste,) the flip-flops and the mustard that, according to Kate, were hoarded by the teenager.

77. It was (the console,) the parka and the pineapple that, according to Charles, were taken by the mugger.

78. It was (the ferret,) the charger and the ham that, according to Kwame, were stolen by the cousin.

79. It was (the sheet,) the clothes and the brisket that, according to Lydia, were supplied by the organization.

80. It was (the paperwork,) the bracelet and the croissant that, according to Albert, were obtained by the refugee.

81. It was (the form,) the tulip and the melon that, according to Nelson, were received by the patient.

82. It was (the bartender,) the atmosphere and the pho that, according to Andrew, were praised by the reviewer.

83. It was (the chessboard,) the mansion and the flask that, according to Mark, were given by the donor.

84. It was (the piano,) the pasta and the beanie that, according to Vincent, were contributed by the trustee.

85. It was (the whisk,) the doghouse and the skirt that, according to Flora, were presented by the salesperson.

86. It was (the gerbil,) the ticket and the belt that, according to Derek, were gifted by the uncle.

87. It was (the flight,) the eggnog and the heels that, according to Olivia, were provided by the boss.

88. It was (the gin,) the bungalow and the jacket that, according to Ava, were preferred by the retiree.

89. It was (the song,) the portfolio and the pants that, according to Mia, were chosen by the committee.

90. It was (the yoyo,) the beach and the stockings that, according to Charlotte, were selected by the orphan.

91. It was (the engine,) the yogurt and the bag that, according to Amelia, were manufactured by the state.

92. It was (the ring,) the motor and the soap that, according to Harper, were crafted by the artisan.

93. It was (the elevator,) the bone and the brush that, according to Evelyn, were broken by the explosion.

94. It was (the chihuahua,) the underwear and the saucer that, according to Abigail, were thrown by the maniac.

95. It was (the steel,) the swordfish and the chocolate that, according to Alexis, were licked by the infant.

96. It was (the amplifier,) the skull and the beer that, according to Emily, were touched by the toddler.

97. It was (the dung,) the lilac and the vodka that, according to Ella, were smelled by the father.

98. It was (the divorce,) the score and the fog that, according to Camille, were predicted by the forecaster.

99. It was (the fruit,) the tricycle and the boulder that, according to Aria, were hurled by the pitcher.

100. It was (the country,) the commander and the gold that, according to Penelope, were targeted by the pirate.

101. It was (the heirloom,) the husky and the jade that, according to Avery, were retrieved by the family.

102. It was (the mayor,) the sculpture and the valley that, according to Mila, were loved by the people.

103. It was (the cuisine,) the tram and the creek that, according to Scarlett, were liked by the locals.

104. It was (the pug,) the peony and the violin that, according to Layla, were transported by the driver.

105. It was (the hall,) the understudy and the saxophone that, according to Chloe, were investigated by the detective.

106. It was (the unicycle,) the branch and the bassoon that, according to Victoria, were hauled by the musician.

107. It was (the limousine,) the alcohol and the sapphire that, according to Eleanor, were returned by the criminal.

108. It was (the alley,) the viper and the zinc that, according to Nora, were noticed by the pedestrian.

109. It was (the tilapia,) the battery and the grape that, according to Lucy, were bagged by the cashier.

110. It was (the crawfish,) the fridge and the fig that, according to Audrey, were procured by the intermediary.

111. It was (the petal,) the herring and the zucchini that, according to Bella, were sliced by the aunt.

112. It was (the mop,) the trout and the watermelon that, according to Brooklyn, were carried by the shopper.

113. It was (the hardwood,) the reindeer and the chainsaw that, according to Savannah, were showcased by the lumberjack.

114. It was (the minivan,) the sauce and the drill that, according to Claire, were brought by the friend.

115. It was (the uniform,) the lollipop and the sharpie that, according to Naomi, were held by the principal.

116. It was (the pickles,) the video and the thread that, according to Caroline, were produced by the seamstress.

117. It was (the dandelion,) the frame and the stool that, according to Gabriella, were moved by the renter.

118. It was (the warehouse,) the meal and the bunk that, according to Samantha, were assembled by the daughter.

119. It was (the homeowner,) the birch and the rack that, according to Hailey, were cursed by the witch.

120. It was (the radish,) the bulldog and the balcony that, according to Delilah, were cleaned by the butler.

121. It was (the culprit,) the ladder and the fire that, according to Eva, were detected by the firefighter.

122. It was (the crate,) the taxi and the pasture that, according to Piper, were tidied by the farmhand.

123. It was (the ranch,) the wife and the brooch that, according to Sadie, were located by the husband.

124. It was (the bandana,) the bouquet and the letter that, according to Josephine, were misplaced by the acquaintance.

125. It was (the yacht,) the escalator and the bumper that, according to Sansa, were fixed by the mechanic.

126. It was (the cardboard,) the blade and the vine that, according to Liam, were bent by the beast.

127. It was (the RV,) the headband and the deodorant that, according to Noah, were used by the influencer.

128. It was (the cinematographer,) the resort and the eyeliner that, according to Oliver, were promoted by the model.

129. It was (the slippers,) the chorizo and the sash that, according to Elijah, were hung by the janitor.

130. It was (the diet,) the peppermint and the floss that, according to William, were recommended by the dentist.

131. It was (the cathedral,) the trolley and the wire that, according to James, were repaired by the electrician.

132. It was (the ravioli,) the marigold and the plug that, according to Benjamin, were kept by the handyman.

133. It was (the vinegar,) the mistletoe and the napkin that, according to Lucas, were ordered by the nephew.

134. It was (the cupboard,) the meadow and the road that, according to Levi, were scoured by the bandits.

135. It was (the raft,) the flag and the cod that, according to Sebastian, were speared by the fisherman.

136. It was (the wizard,) the azalea and the deer that, according to Mateo, were worshipped by the cavepeople.

137. It was (the pop-up,) the setting and the vanilla that, according to Jack, were relished by the expert.

138. It was (the safehouse,) the general and the bison that, according to Owen, were attacked by the smugglers.

139. It was (the officer,) the tax and the wolf that, according to Theodore, were detested by the rancher.

140. It was (the teacup,) the bloom and the porch that, according to Aiden, were rearranged by the gardener.

141. It was (the guideline,) the bra and the torch that, according to Samuel, were forgotten by the guest.

142. It was (the safe,) the manatee and the lighthouse that, according to Joseph, were safeguarded by the patrol.

143. It was (the support,) the porridge and the bomb that, according to Nathan, were solicited by the insurrectionists.

144. It was (the cooperation,) the mutton and the mask that, according to Caleb, were demanded by the authority.

145. It was (the seafood,) the arms and the fur that, according to Santiago, were stockpiled by the aristocrat.

146. It was (the brunch,) the doctor and the cloth that, according to Ezekiel, were fetched by the chauffeur.

147. It was (the petunia,) the supper and the knob that, according to Brooks, were rejected by the locksmith.

148. It was (the contractor,) the site and the brick that, according to Xavier, were evaluated by the builder.

149. It was (the proprietor,) the origami and the pumpkin that, according to Kai, were fancied by the lawyer.

150. It was (the popcorn,) the jersey and the ball that, according to Jose, were grabbed by the referee.

151. It was (the profit,) the light and the show that, according to Riley, were appreciated by the playwright.

152. It was (the ivy,) the microbe and the cocaine that, according to Zoey, were contained by the specialist.

153. It was (the cardiologist,) the syringe and the bacteria that, according to Hannah, were avoided by the nurse.

154. It was (the inheritance,) the puppy and the dinner that, according to Hazel, were shared by the household.

155. It was (the classroom,) the dean and the story that, according to Violet, were complimented by the professor.

156. It was (the fossil,) the sloth and the nail that, according to Stella, were extracted by the paleontologist.

157. It was (the centipede,) the moss and the water that, according to Natalie, were removed by the cleaner.

158. It was (the booklet,) the daffodil and the panna cotta that, according to Emilia, were left by the attendee.

159. It was (the backpack,) the eel and the marmalade that, according to Addison, were boxed by the seller.

160. It was (the crop,) the turkey and the rake that, according to Carolina, were rated by the farmer.

161. It was (the log,) the beret and the pistol that, according to Maya, were swung by the enemy.

162. It was (the nightstand,) the vegetables and the metal that, according to Valentina, were gathered by the vagabond.

163. It was (the furniture,) the sedan and the silk that, according to Ruby, were replaced by the producer.

164. It was (the grain,) the earmuffs and the bandage that, according to Kennedy, were distributed by the government.

165. It was (the institute,) the president and the waterfall that, according to Ivy, were guarded by the ranger.

166. It was (the minister,) the jet and the town that, according to Ariana, were protected by the military.

167. It was (the lioness,) the cottage and the peninsula that, according to Allison, were submerged by the tsunami.

168. It was (the reward,) the suit and the screen that, according to Adeline, were exchanged by the gamer.

169. It was (the raincoat,) the parrot and the tripod that, according to Emery, were auctioned by the celebrity.

170. It was (the service,) the freelancer and the pencil that, according to Lydia, were advertised by the firm.

171. It was (the workshop,) the spruce and the easel that, according to Clara, were exhibited by the painter.

172. It was (the handprint,) the armor and the telescope that, according to Vivian, were displayed by the librarian.

173. It was (the swimsuit,) the flour and the flooring that, according to Peyton, were dried by the heat.

174. It was (the suite,) the convertible and the bucket that, according to Julia, were borrowed by the relative.

175. It was (the companion,) the skyscraper and the puzzle that, according to Arya, were enjoyed by the enthusiast.

176. It was (the flint,) the root and the wheat that, according to Henry, were ground by the villager.

177. It was (the bug,) the pollen and the salt that, according to Alexander, were separated by the peasant.

178. It was (the clothing,) the comforter and the soda that, according to Casey, were dispensed by the warden.

179. It was (the vacation,) the jacuzzi and the honey that, according to Michael, were craved by the runner.

180. It was (the grass,) the quartz and the peanut that, according to Ethan, were crushed by the manufacturer.

181. It was (the grasshopper,) the marijuana and the newspaper that, according to Jacob, were surrendered by the niece.

182. It was (the dictator,) the jewel and the tower that, according to Logan, were defended by the troops.

183. It was (the cedar,) the hippo and the bridge that, according to Jackson, were surrounded by the barriers.

184. It was (the guide,) the friendliness and the cafe that, according to David, were cherished by the traveler.

185. It was (the waste,) the leader and the brewery that, according to Wyatt, were hated by the community.

186. It was (the sequoia,) the party and the inn that, according to Carter, were bombed by the cultists.

187. It was (the leech,) the smog and the tent that, according to Asher, were disliked by the backpackers.

188. It was (the shelf,) the motorbike and the palm that, according to Julian, were relocated by the planner.

189. It was (the entertainment,) the helper and the asparagus that, according to Jayden, were needed by the host.

190. It was (the labrador,) the overalls and the headphones that, according to Ezra, were desired by the girl.

191. It was (the doll,) the cockatoo and the salsa that, according to Joshua, were requested by the boy.

192. It was (the pomegranate,) the bow and the chalk that, according to Silas, were snatched by the teacher.

193. It was (the canary,) the violet and the journal that, according to Amir, were owned by the prisoner.

194. It was (the recliner,) the book and the ale that, according to Enzo, were packed by the porter.

195. It was (the textile,) the time and the cider that, according to Omar, were wasted by the employer.

196. It was (the safari,) the consultant and the subway that, according to Javier, were introduced by the governor.

197. It was (the volunteer,) the facility and the shore that, according to Griffin, were misused by the investor.

198. It was (the apprentice,) the alligator and the poppy that, according to Zain, were treasured by the researcher.

199. It was (the hare,) the civilian and the forest that, according to Ramon, were harmed by the war.

200. It was (the seating,) the keychain and the naan that, according to Remi, were offered by the vendor.

REFERENCES

Anderson, J. R. (1996). Act: A simple theory of complex cognition. *American Psychologist*, 51(4):355.

Arnett, N. and Wagers, M. (2017). Subject encodings and retrieval interference. *Journal of Memory and Language*, 93:22–54.

Avons, S. (1999). Effects of visual similarity on serial report and item recognition. *The Quarterly Journal of Experimental Psychology: Section A*, 52(1):217–240.

Awh, E., Barton, B., and Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, 18(7):622–628.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11):417–423.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, 18(4):362–365.

Baddeley, A. D. and Hitch, G. (1974). Working memory. In *Psychology of Learning and Motivation*, volume 8, pages 47–89. Elsevier.

Barker, J., Nicol, J., and Garrett, M. (2001). Semantic factors in the production of number agreement. *Journal of Psycholinguistic Research*, 30(1):91–114.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bays, P. M. and Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, 321(5890):851–854.

Berggren, N. and Eimer, M. (2016). Does contralateral delay activity reflect working memory storage or the current focus of spatial attention within visual working memory? *Journal of Cognitive Neuroscience*, 28(12):2003–2020.

Brady, T. F. and Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, 120(1):85.

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., and Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural mr images of the brain. *IEEE Transactions on Medical Imaging*, 18(1):32–42.

Bürkner, P.-C. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1):1–28.

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

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., and Haynes, J.-D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience*, 21(4):494–496.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, 55(1):75–84.

Cowan, N. (1999). *An Embedded-Processes Model of Working Memory*, page 62–101. Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19(1):51–57.

Craik, F. I. and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3):268.

Cruz Heredia, A. A., Dickerson, B., and Lau, E. (2022). Towards understanding sustained neural activity across syntactic dependencies. *Neurobiology of Language*, 3(1):87–108.

Cunnings, I. and Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, 75:117–139.

Cunnings, I. and Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language*, 102:16–27.

Curtis, C. E. and D'Esposito, M. (2006). Selection and maintenance of saccade goals in the human frontal eye fields. *Journal of Neurophysiology*, 95(6):3923–3927.

Curtis, C. E. and Sprague, T. C. (2021). Persistent activity during working memory from front to back. *Frontiers in Neural Circuits*, 15:696060.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Diaz, G. K., Vogel, E. K., and Awh, E. (2021). Perceptual grouping reveals distinct roles for sustained slow wave activity and alpha oscillations in working memory. *Journal of Cognitive Neuroscience*, 33(7):1354–1364.

Dillon, B., Clifton Jr, C., and Frazier, L. (2014). Pushed aside: Parentheticals, memory and processing. *Language, Cognition and Neuroscience*, 29(4):483–498.

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

Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19(1):158–164.

Drummond, A. (2013). Ibex farm. *Online server: http://spellout. net/ibexfarm*.

Emrich, S. M., Riggall, A. C., LaRocque, J. J., and Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *Journal of Neuroscience*, 33(15):6516–6523.

Engelmann, F., Jäger, L. A., and Vasishth, S. (2019). The effect of prominence and cue association on retrieval processes: A computational account. *Cognitive Science*, 43(12):e12800.

Fallon, A. B., Groves, K., and Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: When cat helps rat, but not man. *International Journal of Psychology*, 34(5-6):301–307.

Fiebach, C. J., Schlesewsky, M., and Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from erps and fmri. *Journal of Psycholinguistic Research*, 30:321–338.

Fiebach, C. J., Schlesewsky, M., and 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.

Fiebach, C. J., Schlesewsky, M., Lohmann, G., Von Cramon, D. Y., and Friederici, A. D. (2005). Revisiting the role of broca's area in sentence processing: syntactic integration versus syntactic working memory. *Human Brain Mapping*, 24(2):79–91.

Fields, E. (2017). Factorial mass univariate erp toolbox [computer software].

Fields, E. C. and Kuperberg, G. R. (2020). Having your cake and eating it too: Flexibility and power with mass univariate statistics for erp data. *Psychophysiology*, 57(2):e13468.

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., and Awh, E. (2016). The topography of alpha-band activity tracks the content of spatial working memory. *Journal of Neurophysiology*, 115(1):168–177.

Foster, J. J., Vogel, E. K., and Awh, E. (2019). Working memory as persistent neural activity. In *Oxford Handbook of Human Memory*, volume 1. Oxford University Press.

Frazier, L. and d'Arcais, G. B. F. (1989). Filler driven parsing: A study of gap filling in dutch. *Journal of Memory and Language*, 28(3):331–344.

Fukuda, K., Mance, I., and Vogel, E. K. (2015). α power modulation and event-related slow wave provide dissociable correlates of visual working memory. *Journal of Neuroscience*, 35(41):14009–14016.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331–349.

Fuster, J. M. (1995). Memory and planning: Two temporal perspectives of frontal lobe function. *Avdvances in Neurology*, 66:9–20.

Fuster, J. M. and Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*, 173(3997):652–654.

Gao, Z., Gao, Q., Tang, N., Shui, R., and Shen, M. (2016). Organization principles in visual working memory: Evidence from sequential stimulus display. *Cognition*, 146:277–288.

Gardiner, J. M., Craik, F. I., and Birtwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, 11(6):778–783.

Garnsey, S. M. (1986). *Function words and content words: Reaction time and evoked potential measures of word recognition.* PhD thesis, University of Rochester. Department of Psychology.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, 2000:95–126.

Gordon, P. C., Hendrick, R., and Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(6):1411.

Grodner, D. and Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29(2):261–290.

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields i: A critical tutorial review. *Psychophysiology*, 48(12):1711–1725.

Gupta, P., Lipinski, J., and Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, 33:1001–1016.

Hagiwara, H., Soshi, T., Ishihara, M., and Imanaka, K. (2007). A topographical study on the event-related potential correlates of scrambled word order in japanese complex sentences. *Journal of Cognitive Neuroscience*, 19(2):175–193.

Hakim, N., Feldmann-Wüstefeld, T., Awh, E., and Vogel, E. K. (2020). Perturbing neural representations of working memory with task-irrelevant interruption. *Journal of Cognitive Neuroscience*, 32(3):558–569.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, NAACL '01, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

Hanley, J. R. and Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity: A test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review*, 10:435–444.

Harrison, S. A. and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238):632–635.

Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5):393–402.

Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, 26(3):376–405.

Hofmeister, P. and Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology*, 5:1237.

Honnibal, M. and Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.

Huang, L. and Awh, E. (2018). Chunking in working memory via content-free labels. *Scientific Reports*, 8(1):23.

Huttenlocher, J. and Newcombe, N. (1976). Semantic effects on ordered recall. *Journal of Verbal Learning and Verbal Behavior*, 15(4):387–399.

Ishiguro, S. and Saito, S. (2021). The detrimental effect of semantic similarity in short-term memory tasks: A meta-regression approach. *Psychonomic Bulletin & Review*, 28(2):384–408.

Jäger, L. A., Engelmann, F., and Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and bayesian meta-analysis. *Journal of Memory and Language*, 94:316–339.

Jha, A. P. and McCarthy, G. (2000). The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional mri study. *Journal of Cognitive Neuroscience*, 12(Supplement 2):90–105.

Jiang, Y., Chun, M. M., and Olson, I. R. (2004). Perceptual grouping in change detection. *Perception & Psychophysics*, 66:446–453.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., and Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59:193–224.

Just, M. A. and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, 99(1):122.

Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The p600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2):159–201.

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2):175–219.

Karimi, H., Diaz, M., and Ferreira, F. (2019). "a cruel king" is not the same as "a king who is cruel": Modifier position affects how words are encoded and retrieved from memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(11):2010.

Karimi, H., Diaz, M. T., and Wittenberg, E. (2020). Sheer time spent expecting or maintaining a representation facilitates subsequent retrieval during sentence processing. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Karimi, H. and Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, 69(5):1013–1040.

Karimi, H., Fukumura, K., Ferreira, F., and Pickering, M. J. (2014). The effect of noun phrase length on the form of referring expressions. *Memory & Cognition*, 42(6):993–1009.

Karimi, H., Swaab, T. Y., and Ferreira, F. (2018). Electrophysiological evidence for an independent effect of memory retrieval on referential processing. *Journal of Memory and Language*, 102:68–82.

Kim, S. J. and Xiang, M. (2023). Memory encoding/retrieval is sensitive to discourse status: through the lens of pronoun resolution. In *36th Annual Conference on Human Sentence Processing*.

King, J. and Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30(5):580–602.

King, J. W. and Kutas, M. (1995). Who did what and when? using word-and clause-level erps to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7(3):376–395.

Kluender, R. and Kutas, M. (1993). Bridging the gap: Evidence from erps on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5(2):196–214.

Kreither, J., Papaioannou, O., and Luck, S. J. (2022). Active working memory and simple cognitive operations. *Journal of Cognitive Neuroscience*, 34(2):313–331.

Kush, D., Johns, C. L., and Van Dyke, J. A. (2015). Identifying the role of phonology in sentence-level reading. *Journal of Memory and Language*, 79:18–29.

Kutas, M. and Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & cognition*, 11(5):539–550.

Kutas, M. and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947):161–163.

Kwon, N., Kluender, R., Kutas, M., and Polinsky, M. (2013). Subject/object processing asymmetries in korean relative clauses: Evidence from erp data. *Language*, 89(3):537.

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., and Postle, B. R. (2013). Decoding attended information in short-term memory: an eeg study. *Journal of Cognitive Neuroscience*, 25(1):127–142.

Lau, E. (2018). Neural indices of structured sentence representation: State of the art. *Psychology of Learning and Motivation*, 68:117–142.

Lau, E. and Liao, C.-H. (2018). Linguistic structure across time: Erp responses to coordinated and uncoordinated noun phrases. *Language, Cognition and Neuroscience*, 33(5):633–647.

Lefebvre, C., Vachon, F., Grimault, S., Thibault, J., Guimond, S., Peretz, I., Zatorre, R. J., and Jolicœur, P. (2013). Distinct electrophysiological indices of maintenance in auditory and visual short-term memory. *Neuropsychologia*, 51(13):2939–2952.

Leung, H.-C., Gore, J. C., and Goldman-Rakic, P. S. (2002). Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *Journal of Cognitive Neuroscience*, 14(4):659–671.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Lewandowsky, S., Oberauer, K., and Brown, G. D. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, 13(3):120–126.

Lewis, R. L. and 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., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10):447–454.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., and Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*, 24(1):61–79.

Lin, P.-H. and Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition*, 17(3):356–372.

Lo, C.-W. and Brennan, J. R. (2021). Eeg correlates of long-distance dependency formation in mandarin wh-questions. *Frontiers in Human Neuroscience*, 15:591613.

Logie, R. H., Del Sala, S., Wynn, V., and Baddeley, A. D. (2000). Visual similarity effects in immediate verbal serial recall. *The Quarterly Journal of Experimental Psychology Section A*, 53(3):626–646.

Lopez-Calderon, J. and Luck, S. J. (2014). Erplab: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8:213.

Love, T. and Swinney, D. (1996). Coreference processing and levels of analysis in object-relative constructions; demonstration of antecedent reactivation with the cross-modal priming paradigm. *Journal of Psycholinguistic Research*, 25:5–24.

Lowder, M. W. and Gordon, P. C. (2014). Effects of animacy and noun-phrase relatedness on the processing of complex sentences. *Memory & Cognition*, 42:794–805.

Luck, S. J. and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657):279–281.

Malik-Moraleda, S., Ayyash, D., Gallée, J., Affourtit, J., Hoffmann, M., Mineroff, Z., Jouravlev, O., and Fedorenko, E. (2022). An investigation across 45 languages and 12 language families reveals a universal language network. *Nature Neuroscience*, 25(8):1014–1019.

Mate, J. and Baqués, J. (2009). Visual similarity at encoding and retrieval in an item recognition task. *The Quarterly Journal of Experimental Psychology*, 62(7):1277–1284.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language*, 38(2):225–252.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3):817.

McElree, B. and 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.

Mertzen, D., Laurinavichyute, A., Dillon, B., Engbert, R., and Vasishth, S. (2020). A cross-linguistic investigation of proactive, similarity-based retrieval interference in sentence comprehension: No support from english, german and russian eye-tracking data. *PsyArXiv*.

Mertzen, D., Paape, D., Dillon, B., Engbert, R., Vasishth, S., Mertzen, D., Paape, D., Dillon, B. W., Engbert, R., and Vasishth, S. (2023). Syntactic and semantic interference in sentence comprehension: Support from english and german eye-tracking data. *Glossa Psycholinguistics*, 2(1).

Münte, T. F., Schiltz, K., and Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395(6697):71–73.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18(3):251–269.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, 10(5-6):389–395.

Ness, T. and Meltzer-Asscher, A. (2017). Working memory in the processing of long-distance dependencies: Interference and filler maintenance. *Journal of Psycholinguistic Research*, 46:1353–1365.

Ness, T. and Meltzer-Asscher, A. (2019). When is the verb a potential gap site? the influence of filler maintenance on the active search for a gap. *Language, Cognition and Neuroscience*, 34(7):936–948.

Nicenboim, B., Vasishth, S., Engelmann, F., and Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in german. *Cognitive Science*, 42:1075–1100.

Nicol, J. and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18:5–19.

Nolden, S., Grimault, S., Guimond, S., Lefebvre, C., Bermudez, P., and Jolicoeur, P. (2013). The retention of simultaneous tones in auditory short-term memory: a magnetoencephalography study. *NeuroImage*, 82:384–392.

Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3):411.

Oberauer, K. (2019). Working memory and attention — a conceptual analysis and review. *Journal of Cognition*.

Oberauer, K. and Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55(4):601–626.

Phillips, C., Kazanina, N., and Abada, S. H. (2005). Erp effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22(3):407–428.

Politzer-Ahles, S., Xiang, M., and Almeida, D. (2017). “before" and “after": Investigating the relationship between temporal connectives and chronological ordering using event-related potentials. *PloS One*, 12(4):e0175199.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, 139(1):23–38.

Postle, B. R. (2015). The cognitive neuroscience of visual short-term memory. *Current Opinion in Behavioral Sciences*, 1:40–46.

Rich, S. and Wagers, M. (2020). Semantic similarity and temporal contiguity in subject verb dependency processing. In *33rd Annual CUNY Conference on Human Sentence Processing*, page 11.

Riggall, A. C. and Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *Journal of Neuroscience*, 32(38):12990–12998.

Ristic, B., Mancini, S., Molinaro, N., and Staub, A. (2021). Maintenance cost in the processing of subject–verb dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., and Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, 354(6316):1136–1139.

Saito, S., Logie, R. H., Morita, A., and Law, A. (2008). Visual and phonological similarity effects in verbal immediate serial recall: A test with kanji materials. *Journal of Memory and Language*, 59(1):1–17.

Schrum, N. and Sprouse, J. (2022). The sustained anterior negativity (san) as a diagnostic for movement in how come questions. In *The 58th Annual Conference of the Chicago Linguistic Society*.

Serences, J. T., Ester, E. F., Vogel, E. K., and Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, 20(2):207–214.

Smith, G. and Vasishth, S. (2020). A principled approach to feature selection in models of sentence processing. *Cognitive Science*, 44(12):e12918.

Smyth, M. M., Hay, D. C., Hitch, G. J., and Horton, N. J. (2005). Serial position memory in the visual—spatial domain: Reconstructing sequences of unfamiliar faces. *The Quarterly Journal of Experimental Psychology Section A*, 58(5):909–930.

Sprague, T. C., Ester, E. F., and Serences, J. T. (2016). Restoring latent visual working memory representations in human cortex. *Neuron*, 91(3):694–707.

Sprouse, J., Kucerova, I., Park, J., Cerrone, P., Schrum, N., and Lau, E. (2021). Sustained anterior negativities and movement dependencies in syntax [manuscript in preparation]. *Department of Linguistics, University of Connecticut.*

Srimal, R. and Curtis, C. E. (2008). Persistent neural activity during the maintenance of spatial position in working memory. *Neuroimage*, 39(1):455–468.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116(1):71–86.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, 86(1):142–164.

Steinhauer, K. and Drury, J. E. (2012). On the early left-anterior negativity (elan) in syntax studies. *Brain and Language*, 120(2):135–162.

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

Stokes, M. G. (2015). 'activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences*, 19(7):394–405.

Stowe, L. A. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1(3):227–245.

Tanenhaus, M. K., Carlson, G. N., and Seidenberg, M. S. (1985). Do listeners compute linguistic representations. *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*, pages 359–408.

Thyer, W., Adam, K. C., Diaz, G. K., Velazquez Sanchez, I. N., Vogel, E. K., and Awh, E. (2022). Storage in visual working memory recruits a content-independent pointer system. *Psychological Science*, 33(10):1680–1694.

Traxler, M. J., Morris, R. K., and Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47(1):69–90.

Troyer, M., Hofmeister, P., and Kutas, M. (2016). Elaboration over a discourse facilitates retrieval in sentence processing. *Frontiers in Psychology*, 7:374.

Tung, T.-Y. and Brennan, J. (2023). Modeling retrieval interference during naturalistic comprehension. In *36th Annual Conference on Human Sentence Processing*.

Ueno, M. and Garnsey, S. M. (2008). An erp study of the processing of subject and object relative clauses in japanese. *Language and Cognitive Processes*, 23(5):646–688.

Unsworth, N., Fukuda, K., Awh, E., and Vogel, E. K. (2015). Working memory delay activity predicts individual differences in cognitive abilities. *Journal of Cognitive Neuroscience*, 27(5):853–865.

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.

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

Villata, S., Tabor, W., and Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, 9:2.

Vogel, E. K. and Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984):748–751.

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

Wagers, M. W. and Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology*, 67(7):1274–1304.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 77(1):1.

Wilken, P. and Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12):11–11.

Wolff, M. J., Ding, J., Myers, N. E., and Stokes, M. G. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience*, 9:123.

Wolff, M. J., Jochim, J., Akyürek, E. G., and Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, 20(6):864–871.

Woodman, G. F., Vecera, S. P., and Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, 10(1):80–87.

Xu, Y. and Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences*, 104(47):18766–18771.

Yang, C. L., Perfetti, C. A., and Liu, Y. (2010). Sentence integration processes: An erp study of chinese sentence comprehension with relative clauses. *Brain and Language*, 112(2):85–100.

Yano, M. and Koizumi, M. (2018). Processing of non-canonical word orders in (in) felicitous contexts: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, 33(10):1340–1354.

Yano, M. and Koizumi, M. (2021). The role of discourse in long-distance dependency formation. *Language, Cognition and Neuroscience*, 36(6):711–729.

Yu, X., Li, J., Zhu, H., Tian, X., and Lau, E. (2023). Electrophysiological hallmarks for event relations and event roles in working memory. *bioRxiv*, pages 2023–05.