THE UNIVERSITY OF CHICAGO

MEMORY ENCODING AND RETRIEVAL OF DISCOURSE STRUCTURE IN HUMAN LANGUAGE COMPREHENSION

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
SANGHEE KIM

CHICAGO, ILLINOIS
AUGUST 2024

To my parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This work and my PhD journey would have never come to an end without the support and guidance I received from so many people.

First and foremost, I would like to express my sincere gratitude to my advisor, Ming Xiang, and committee members, Chris Kennedy, Allyson Ettinger, and Monica Do.

There wasn't a single moment in the past 6 years where Ming was (or at least looked) bothered by my questions or emails. She would spare extra time for meetings, at times as long as even three hours, to help me refine my thoughts and research directions. She asks sharp questions and sometimes I would get nervous before meetings, but I truly loved those question-and-answering periods. She taught me to conduct rigorous science, to discover new findings from experimental data, and most importantly, to enjoy the process. She also guided me to present myself professionally. Aside from research, Ming never forgot to ask how "Sanghee" was doing. Thank you, Ming, it's been a true joy learning from you and working with you.

Chris encouraged me to think about the "big(ger) question." He also assured me that I was making good progress and making contributions to the field—and that felt quite good. Chris was one of the readers of my first qualifying paper. That was a difficult period for me, and Chris helped me by advising me to focus on completing each task at hand, day by day. I passed my first QP, and this experience gave me a life lesson that taking baby steps and focusing on the things right in front of me is the key to navigating the way out when I'm stuck. Thank you, Chris, for the encouragement and guidance.

I consider myself lucky to have met Allyson during my PhD. Allyson gave me the courage to dabble in the computational linguistics world, and she introduced me to the joy of coding and reading NLP-related papers. Her lab meetings and course discussions were some of my favorite times, with lots of brain-tickling moments. Working with her for two years, I not only gained knowledge about compling but also learned to work efficiently and speak and

write clearly. Thank you, Allyson, for all your wise advice, support, and kindness.

Monica is the person to go to when you're stuck on "packaging." Whether it's writing a paper or dissertation, presenting a talk, preparing for an interview, or teaching, ask Monica for her help and she will do magic. Even when I bring a salad bowl of thoughts, disorganized, she would play the role of a librarian, placing them in the right place. Monica shared with me so many resources that were crucial for completing this project, too. Thank you, Monica.

Prior to diving into my dissertation project, I also had the privilege of working with and being advised by other faculty in the department: Karlos Arregi, Anastasia Giannakidou, and Diane Brentari. I still remember the meetings with Karlos during my first year. Anyone who has met Karlos would be familiar with his excitement for morphology and syntax. I'm grateful to have had the opportunity to meet regularly with him and be part of those thrilling moments. Anastasia was the other reader for my first QP. She encouraged me to continue developing the final paper I wrote for her class. I thank her for kindly walking me through the literature and showing me how to think like a semanticist. I didn't become one, but it was a great learning experience that taught me to read semantics papers. Finally, I was luckily part of a sign language project, and Diane was on our team. Her intuition was always right and made sense. It was great to have worked with her and have experienced those insightful moments—thank you, Diane.

I also thank the faculty in the Linguistics department, especially, Alan Yu, Itamar Francez, and Erik Zyman. They were always kind and supportive whenever I reached out for help or advice. Alan and Itamar kindly offered help when I was recovering from a car accident, connecting me to lawyers and buying me a nice warm hot chocolate. Erik and I were Franke Institute Fellows in the same year, and it was a wonderful learning experience to see him communicate his expertise in linguistics to non-linguists. Erik generously made time to help me with research. Thank you, all.

My gratitude also goes to the professors I met when I was still an undergrad or master's

student: Mi-Jeong Song, Jeong-Ah Shin, Chang-Yong Sohn, Ki-Sun Hong, Heejeong Ko, Nayoung Kwon, Taehong Cho, Miseon Lee, Sun-Ah Jun, Sung-Ho Ahn, and Munwoo Lee. Either taking their class or working with them inspired me to want to be in academia, and they have been all so supportive of my goals.

I am grateful to have met my dear academic siblings (aka. Ming's minions): Daniel Lam, Eszter Ronai, and (the honorary member) Lucas Fagen. Daniel and I share many research interests and have similar backgrounds. He would play the role of a tutor answering all my questions, be they experimental methods or statistical analyses. Daniel is one of the most warmhearted people I know, and he will always brighten your perspective. Thank you, Daniel, for the jokes and for reminding me not to take things too seriously. Eszter was my guidebook. She shared her experiences and pro-tips helping me with fellowships, job applications, and class preparations. Spending time with her energizes me—I like her energy and confidence, not to mention her sense of humor. Thank you for your encouragement and the delicious chocolate cookies, Eszter. Lucas and I have been co-coordinators for the LEAP workshop for two years. Lucas is a fantastic person to work with, great at communication, and responsible. He is a great listener and friend, too. Thank you, Lucas.

I thank my fellow CLS officers, Corinne Kasper, Matt Hewett, Naomi Kurtz, and my special cohort member, Zeineb Sellami. We went through the COVID era together, and I think that created some sense of camaraderie between us. I also thank my sign language project crew, Aurora Martinez del Rio, Casey Ferrara, and Emre Hakgüder. We worked hard and learned a lot, but also had a lot of fun playing video games and making pasta. Yenan Sun, Jackie Lai, and Dan Edmiston were the very first people who convinced me to join UChicago during the open house visit. I'm especially thankful to Yenan, who, as everybody knows, is thoughtful and kind. I owe her all the hours that she spent giving me tours of Hyde Park, sharing good food, and walking in the park together. I thank Thomas Sostarics for hours and hours of chit-chatting. He also deserves a special shout-out for helping me

with the experimental setup. Thank you for responding to me even when I bombarded you with so many questions. I thank Laura Stigliano for inviting me to watch the Argentina vs. Australia soccer match during the 2022 World Cup. I thank my swimming buddies, Brianna Wilson, and Serpil Karabüklü, for the healthy Saturdays. I thank Akshay Aitha for the fun chats about our dating philosophy. I thank Mike Tabatowski for the tasty brunch and cakes. I thank Jaehong Shim and Xiaobei Chen for the opportunities to talk in and about Korean. I also thank Saulé Tuganbaeva for her help when I was visiting Moscow. I also thank my fellow TA-s, Ömer Eren, Pamela Sugrue, and Kat Montemurro, the instructors I TA-ed for, Jacob Phillips, and Patrick Munõz, and fellow graduate students, Steven Castro, Mina Giannoula, Sam Gray, Kate Henninger, Kutay Serova, and others.

I appreciate the support I had from the labs. I thank the members of the Language Processing Lab, especially, Jinghua Ou, Parker Robbins, Tzu-Yun Tung, and Daniel Jiayuan Yue. I thank the members of the Computational Linguistics Lab, especially, Kanishka Misra—he was my encyclopedia when I first got into compling—and Lang Yu, and Lalchand Pandia—my collaborators—and Freda Shi, Chih-chan Tien, and Jiaxuan Li. I'm especially grateful to Mourad Heddaya, who baked cakes and pies for me whenever there was anything to celebrate, and made the most delicious honey lemon ginger tea when I needed comfort. He unlocked the brave and adventurous side of me, and I've never looked back since. Thank you for making me laugh, and introducing me to Curb Your Enthusiasm.

I cannot forget to thank my non-linguist friends in Hyde Park. Youjin Lee was my friend, sister, aunt, mom, whatever you name it. I am forever indebted to her for her kindness. She literally fed me and helped me fall asleep. I wouldn't have been able to survive my first year in Chicago without her. Sangmin Oh was the biggest cheerleader. He lifted me up to make me walk again when I fell. We share lots of fond memories together, which I dearly cherish as part of my PhD journey. I thank Hwanwoo Kim, Injae Kim, and Kiho Park, who I consider my Chicago brothers. Dong-Hyuk Seo, Jaehyeok Jin, and Yewon Kim have been

an inspiration academically, and I thank them for their kindness. I also thank Diane Oh, and Flora Zhang, for the good food and coffee. I am grateful to have met Chang Liu, Joohee Cho, and Preethika Thiraviam, my wonderful housemates. I thank Melissa Weber for the amazing pottery classes, and Kayla Haggart—then an instructor, now a good friend—for her serene yoga and top-roping climbing lessons (with Max). Thanks, everyone.

I also thank my Korean (ex-)linguist friends. Jinyoung Jo, Juyeon Cho, and I have been on the same boat since 2015. We've been in different states, but laughed and cried over everything related to linguistics, relationships, and family matters. I cannot thank them enough for their encouragement and kind words—thank you, my "bamboo trees." And to my "New York" friends, Semi Park, and Hyunjung Joo: thank you for the fun phone calls. Giggling and laughing about mundane things with you always lightened up my day. I was also lucky to have my fellow neuro/psycholinguist, Sarah Hye-yeon Lee, and Jeongwha Cho. I also thank Heemin Park for her fun stories about making makgeolli.

Thank you to my friends in Korea, who supported and encouraged me. Seoyeon Kwon already called me Dr. Kim even before I started my PhD program. But now I get to reveal the news that she can officially call me Dr. Kim. Soojung Kim, Suhyun Son, and Heeju Yoo, always made sure I had tasty Korean food and updated me with the most recent K-culture. I thank Dabom Kim, Hyaewon Lee, Hyejin Cho, Seohyun Chung, and Sungyon Hong—the trips with them always filled me with the energy to return to Chicago and roll. I thank Hyeon-Rae Cho (and Sungyon) for waking up at 3 AM to hike and catch the sunrise with me. And Hyunjin Kim, and Yeojin Kim: dilly-dallying on your couch, cooking outdoors on an ocean drive, and madly clapping to make fish jump in the lake were the memories that I held onto whenever I missed Korea. Thank you, all.

And to my family. My parents have been the true inspiration for me choosing the path of pursuing a PhD. My dad is someone who finds wonder in everything around us. One of his life mottos is to "learn and share, and learn to share," and I believe this would give you

some idea of who he is. And my mom—we say that her middle name is *steadiness*. Once she decides to do something, you can find her doing it for years without skipping a day. Perhaps I inherited both of their traits and that made it possible for me to complete the 6-year-long PhD program. I am also lucky to have such loving parents, who believed in me more than I did myself. I also thank Stan *oppa*, who has been such a good friend and brother. He is one of the wisest people I know. He assured me that there are no wrong decisions—I just need to make one, get the most out of it and learn from its outcome. I thank Sonya, my sister-in-law, for her kindness and encouragement. I also thank my grandma, and aunt (*komo*) in LA, who gave me lots of advice on living an independent life.

The project was supported by the generous NSF DDRI Grant BCS-2214437. I thank the grant reviewers for finding value in this work. I am grateful to the Franke Institute for the dissertation fellowship and to the fellows for their thoughtful comments on the first draft of the chapters. I thank the reviewers and audience at Human Sentence Processing 2022 & 2023, the Annual Meeting of the Linguistic Society of America, for their comments, and the Gibson-Fedorenko Award committee. I especially thank Jennifer Arnold, Brian Dillon, and Matt Wagers for their suggestions at the early stage of this project, and Jérémy Zehr for his technical support. I am also grateful to Tal Linzen, and Marten van Schijndel, for their mentorship. I had the opportunity to defend my dissertation at the UChicago Center in France. The defense happened during my research visit at Laboratoire de Linguistique Formelle in Paris. This was possible with the support of Caterina Donati and Barbara Hemforth—thank you for such an amazing opportunity. I thank everyone who joined my defense, both from far away and in person, making the event even more special.

ABSTRACT

An extensive amount of work has shown the way the working memory architecture supports language comprehension. However, while language comprehension inevitably takes place in a discourse context, less is known about how information related to discourse is organized and managed in working memory. The project particularly focuses on two known processes in the working memory involved in language comprehension, namely *encoding* and *retrieval*. I investigate the way discourse structure information guides retrieval and how the information affects the encoding process in real-time comprehension.

To this end, I use two types of sentence structures that are minimally different in their discourse status. While both sentences in (1) contain the information that *the waitress* is "Kelly's sister," the discourse status of such information differs when it is contained in a restrictive relative clause (RRC) (1a) or an appositive relative clause (ARC) (1b). While the content inside an RRC (1a) is part of primary, essential information (i.e., main discourse information), it becomes secondary, and side-commentary (i.e., subordinate discourse information) when it is contained inside an ARC (1b).

(1) a. The waitress *who is Kelly's sister* sat next to Bob.

b. The waitress, *who is Kelly's sister*, sat next to Bob.

Given the contrast of main vs. subordinate discourse information status, I examine how this division affects the retrieval process and what effects this information results in during the encoding process. To address these questions, I make use of two of the well-established linguistic phenomena in the sentence processing literature to examine the contrast: number agreement attraction effect, and pronoun resolution. I conduct a series of experiments using behavioral measures such as reading times using a self-paced reading task and eye-gaze times implementing a visual world paradigm.

The findings suggest that discourse structure information imposes a constraint on the

memory retrieval process. Specifically, *the active state of discourse question* in the given discourse structure guides which linguistic entities can be targeted for retrieval. Furthermore, the results show that discourse structure information affects retrieval even when it is not necessarily used for governing the grammaticality of linguistic dependency. While the distinction in discourse status does not affect grammaticality for resolving the dependency, it impacts the ease of dependency resolution in real time. The results also suggest that the parser actively uses discourse structure information during the encoding process, to the extent that the overlap of this information between encoded linguistic representations leads to a competition, encoding interference effect.

CHAPTER 1

INTRODUCTION

Language functions as a fundamental tool for human interaction and communication, whether expressed in writing, spoken words, or signs (as in sign languages). The comprehension of language appears to occur effortlessly, with the human mind handling it seamlessly. However, this ease does not imply that language understanding is always simple. Challenges arise in situations such as reading a newspaper, or listening to a podcast that contains lengthy and complicated sentences or words that are complex or confusing to keep track of. This dissertation project delves into these aspects of language understanding, examining moments when comprehension difficulties emerge and investigating the reasons behind them. It is important to note that the primary goal of this work is not to simply identify the instances of *failing* moments in human comprehension; rather, those *failures* serve as a window through which we gain insights about the cognitive mechanism that underlies human language comprehension.

One of the crucial observations that was made by linguists was that people tend to find sentences more difficult to comprehend, sometimes even appearing grammaticality unacceptable, as the sentences grow longer and more complex. An illustration of this observation is seen in the incremental drop in comprehensibility from (1a) to (1c). The difficulty in understanding more complicated sentences spurred the idea that language comprehension is tightly linked and supported by the architecture of working memory. Given the limited working memory capacity of humans, it is natural to expect that comprehending lengthy sentences becomes cognitively burdensome, requiring the retention and tracking of more information.

(1) a. The administrator lost the medical reports.

b. The administrator who the intern had bothered lost the medical reports.

c. The administrator who the intern who the nurse supervised had bothered lost the medical reports. (Gibson, 1998)

However, experimental findings reveal that the complexity of a sentence goes beyond its absolute length. For example, despite the same number of words in (2a) and (2b), a structure like (2a) has been reported to be easier to comprehend than the one like (2b). Proposals have been made about the different factors that affect comprehensibility such as the syntactic complexity or the frequency of structures, or the number of meaningful discourse entities (Gibson, 1998, 2000).

(2) King and Just (1991)

a. The reporter who attacked the senator admitted the error.

b. The reporter who the senator attacked admitted the error.

The intricate nature of cognitive processes that underlie language comprehension is exhibited even when people comprehend ungrammatical sentences. Both sentences in (3) are ungrammatical (an asterisk(*) denotes ungrammatical/unacceptable) as *praise* is an incorrect form of the verb that corresponds—*agrees*—with the subject, *the musician*. Ungrammatical sentences generally take longer to read than grammatical sentences, typically at the point where the error occurs. In the example below, reading slows down (i.e., reading time increases) around the word, *praise*, the source of ungrammaticality. However, an important observation from experimental work is that there is a difference in the reading pattern, where reading times are longer in (3a) compared to (3b).

(3) Wagers, Lau, and Phillips (2009)

a. *The musician who the reviewer praise so highly will win a Grammy.

b. *The musician who the reviewers praise so highly will win a Grammy.

This phenomenon has given support to a certain theory about the working memory architecture, which explains that features of linguistic entities are used for dependency resolution. Specifically, as the parser needs to resolve the subject-verb dependency (i.e., subject-verb agreement) at *praise*, direct access to the matching entities in the plural feature (since *praise* is used for a 3rd person plural noun) is initiated. Since *the reviewers* in (3b) has such a feature (i.e., a 3rd person plural noun), the ungrammaticality of (3b) is mitigated, creating a "grammatical illusion" (Phillips, Wagers, & Lau, 2011).

The findings collectively have fostered an understanding of human language comprehension, indicating a strong association and support from working memory, and the organized nature of the working memory architecture. Theories in psycholinguistics have developed in a way to elucidate the principled mechanism, and such an effort has been made by examining various sentential structures and different languages and using behavioral as well as neurological measures to conduct human experiments.

This dissertation builds upon the existing findings on the relation between working memory and language comprehension. While a large body of work has examined the role of semantic, and morphosyntactic representations affecting sentence processing, factors related to discourse have received limited attention. This work seeks to address this gap and focuses on the way that information related to discourse structure guides language comprehension. Specifically, it examines how the division of what is more vs. less important information in discourse affects the way language is comprehended. Would information that is considered important and essential in discourse always be more salient and easy to retrieve from memory? Would information that is less relevant and less significant to the discourse topic be more volatile and have a weaker representation in memory? These questions are addressed in this dissertation by investigations through a series of controlled experiments using different types of linguistic dependencies.

1.1 Retrieving and encoding discourse structure information

A leading theory in the current psycholinguistics literature, sentence processing literature in particular, that accounts for the cognitive architecture underlying human language comprehension is known as the CUE-BASED RETRIEVAL THEORY. The core idea of this theory is that language comprehension, especially sentence comprehension, involves ENCODING and RETRIEVAL of linguistic items and representations, and successful comprehension is carried out by the working memory architecture using certain linguistic FEATURES as CUES to retrieve the previously stored linguistic entities. Focusing on the properties of discourse, I examine (i) the way discourse structure information guides **retrieval**, and (ii) what effects are found during **encoding** discourse structure information.

A large body of work has shown that morphosyntactic and semantic information serve as features and cues to guide memory encoding and retrieval. For instance, in comprehending a sentence as "The musician who the reviewer praises so highly will win a Grammy," the parser encodes the incoming linguistic entities with certain feature representations. *The musician* can be represented with features as [+animate, +singular] and *the reviewers* as [+animate, +plural]. Then encountering *praise* triggers retrieval of relevant linguistic entities that have been previously encoded. To access the relevant entities, previously used features such as [+animate] or [+plural] will be used as access cues to activate and retrieve the relevant linguistic entities. If [+plural] is used as a retrieval cue, entities that were represented with a [+plural] feature will be activated and will most likely be retrieved. These features are commonly represented in a static, binary ([+/−FEATURE]) way. The features would remain constant at any point of parsing the sentence—static—and can easily be expressed in a binary term. For example, *the musician* can be represented with features as [+animate, +singular] at any point of processing the sentence, and *a Grammy* can be expressed with a [−animate] feature.

The first question However, two noticeable properties arise when discourse information is considered. The first pertains to the *dynamic nature of discourse*. Unlike morphosyntactic or semantic information, which can be expressed in a static way, discourse information can have different properties depending on when it is evaluated in the given discourse. Given such dynamicity, how can we understand discourse structure information as a retrieval cue?

Discourse is *dynamic*, as opposed to *static*, with its characteristics shifting throughout the progression of discourse. Such dynamicity is well illustrated in the contrast between the two examples in (4). While the discourse develops in the same way until S_2, it diverges in different ways in S_3. More importantly, the discourse status of the content in S_2 changes as the discourse progresses in different directions. In both cases, the discourse begins by discussing the Millers buying a house. Then the information in S_2 provides further information about S_1, an *explanation* on why the Millers bought the house. In (4a), as the discourse progresses, information in S_3 provides further information about house prices. Since the focus of the discourse is still about house prices, S_2 remains relevant to the most focused topic in the discourse even when the discourse has progressed to the next content. In (4b), on the other hand, S_3 provides information about a separate event other than the house price. Now the topic is (more) about the Millers renting out the place. With the shift of discourse, it consequently makes S_2 less relevant to the currently focused topic. Hence, even when S_2 was a relevant discourse unit at the point it was added to the discourse, it is no longer so as the discourse developed with the addition of S_3 in (4b). This illustrates the dynamic nature of discourse, where discourse relations cannot be defined or captured in static terms.

(4) Jasinskaja (2016)

a. S_1: The Millers bought a house in the country.
   S_2: Country house prices started to rise again.
   S_3: They rose by 1.7 %.

b. S_1: The Millers bought a house in the country.

S_2: Country house prices started to rise again.

S_3: They rented it out.

One can make an attempt to represent discourse structure information in by using the binary feature inventory. For example, to express the different *flavor* of discourse units, where some contain secondary, supplementary information whereas others carry primary, essential information, the first part of the content could be expressed roughly as [+primary] and the less important information as [+secondary]. Or if we construe discourse as a hierarchically structured representation, wherein some units are superordinate to others (and these are subordinate to the prior ones), discourse units could perhaps be expressed as [+main] or [+subordinate].[1]

However, this approach encounters immediate challenges. Representing discourse with static and discrete features becomes problematic as it fails to capture this dynamicity. What was considered primary information moments ago can become secondary as the discourse shifts topics, as illustrated in (4). In a binary term, what was labeled as [+primary] can later change to [+secondary], and possibly even vice versa, as the discourse progresses in different directions. This raises the first question of the dissertation: how is discourse structure information used for retrieval given the dynamic nature of discourse?

The second question The second important aspect of discourse is that discourse structure information *does not determine grammaticality*. This contrasts with morphosyntactic information, which is used as retrieval cues due to grammatical requirements and constraints. However, not being crucial information for grammatical dependency does not mean that it is

1. I am not committed to the notion or the name of the label in [+main] or [+subordinate]. Neither does the usage of this have implications on the syntactic relationship between discourse units (e.g., some are syntactically subordinate to the other). The usage of these terms, *main* and *subordinate*, is adopted as an easy way to juxtapose the contrasting ideas of static vs. dynamic representation of discourse—[+main]/[+subordinate] are used to represent the static, binary feature-based approach.

unnecessary information that will be disregarded during processing. In resolving pronouns, for example, saliency or topichood, which is relevant discourse but not necessarily grammatical constraints, becomes influential factors in determining which antecedent is most plausible. Hence, it is an open question whether discourse structure-related information would guide retrieval in the same way as morphosyntactic information—this serves as the second question in the dissertation.

The third question These properties of discourse information and the questions about retrieval lead to the third question about memory encoding. While the retrieval process during language comprehension has been extensively studied, there is less work that has focused on the encoding process. Even within the investigation on memory encoding during language comprehension, the focus has been primarily on examining the role of morphosyntactic features. Given this gap, the dissertation investigates the effects that arise during memory encoding of discourse structure information.

1.2 An overview of the current work

The dissertation addresses the three aforementioned questions to understand the way discourse information guides memory encoding and memory retrieval during language comprehension. The questions are repeated below:

1. How does discourse structure information guide retrieval, given that discourse is dynamic and cannot be represented in a binary, static way?

2. How does discourse structure information guide retrieval, given that discourse information does not determine grammaticality?

3. How does discourse structure information have an effect on memory encoding?

In addressing these questions, I focus on the aspect of *discourse structure* that it is hierarchically structured in terms of the significance of the discourse components. For example, imagine Sally went to a famous restaurant last night, and her friend Tom asked her whether she liked the restaurant. Sally can simply answer, "The food was fantastic." But she could also provide additional information such as "The asparagus soup had great flavor and was well seasoned." She may even digress a little and add less relevant information such as "The plates and utensils were exotic and interesting." These additional pieces of information are supplementary information and are subordinate to the main idea that the food was fantastic. The current work focuses on this division of what discourse component is more important and what is less so in representing discourse structure.

To examine this division at a sentence level, I make use of a sentential structure known as restrictive relative clauses (RRCs) and appositive relative clauses (ARCs). These structures are particularly useful as they enable the testing of contrasting discourse features at a sentence level. For instance, consider the sentences in (5), which embed two different relative clauses (RCs). Both sentences convey similar meanings. However, the content "(the waitress is) Kelly's sister" in the two sentences has a different *flavor* that contributes to the sentential meaning and discourse effect in distinct ways. In the case of (5a), the phrase "who is Kelly's sister" is crucial information, especially in scenarios with multiple waitresses, as it identifies a specific referent. The content within the RRC is essential and primary information, forming part of the main point of an utterance, and thus, contributes to the MAIN DISCOURSE. This contrasts with an ARC, where the information "(the waitress is) Kelly's sister" is side-commentary, secondary, and non-essential to the main content that the waitress recently got married in (5b). Therefore, ARC content is construed to be part of SUBORDINATE DISCOURSE information. The RRC and ARC structures will be used as a pair in experiments to maintain the key contrast between main and subordinate discourse information.

(5) The key contrast of RRC (main discourse) and ARC (subordinate discourse)

a. The waitress *who is Kelly's sister* recently got married. [RRC]

b. The waitress, *who is Kelly's sister*, recently got married. [ARC]

To examine the way such division in discourse guides language comprehension in real time, I make use of two of the known linguistic dependencies, namely (i) subject-verb dependency, and (ii) anaphoric dependency, specifically involving pronoun resolution.

For the subject-verb dependency, a widely studied phenomenon known as the NUMBER AGREEMENT ATTRACTION EFFECT is used (as previously seen in (3)). Previous work has shown that *the musicians* and *the reviewer*, which are both part of the main discourse, lead to an interference effect during memory retrieval. This known case will be compared to the newly tested construction of (6b), exploring whether the same attraction effect will emerge in it. The presence of an attraction effect in the ARC condition will be taken as evidence to suggest that the main discourse unit (*the musicians*) as well as the subordinate discourse unit (*the reviewer*) is a target for retrieval. Three experiments (Experiments 1–3) using the number agreement attraction effect will be conducted, and these address the first question regarding retrieval.

(6) a. *The musicians who the reviewer praise so highly will win a Grammy. [RRC]

b. *The musicians, who the reviewer praise so highly, will win a Grammy. [ARC]

The second type of structure involves a linguistic dependency that involves PRONOUN RESOLUTION (7). What is unique about this anaphoric dependency is that while the antecedent of the pronoun is ambiguous in both clause types (either *the singers* or *the violinists* is the antecedent of *their*), the type of discourse structure that the antecedents are hosted in is different. While the two antecedents, i.e., target for retrieval, are part of the same discourse structure in (7a), they are in distinct discourse structure units in (7b). In (7a), the

plausible antecedents are both part of the main discourse whereas in (7b), one is in the main discourse and the other is part of a subordinate discourse structure. Crucially, the type of discourse structure does not govern the grammaticality that constrains the antecedent of the pronoun. It is thus a question of whether discourse information will nonetheless lead to any differences in resolving the pronoun. A total of three experiments (Experiments 4–6) will be conducted using a similar structure as in (7), and these pertain to addressing the second research question.

(7) a. The singers who invited the violinists invited their mentors to the party. [RRC]

b. The singers, who invited the violinists, invited their mentors to the party. [ARC]

The final set of experiments (Experiments 7–8) also examines pronoun resolution, using a construction identical to that in (7). These experiments address the question about memory encoding, exploring the way discourse structure information impacts processes involved during the encoding stage. Specifically, would the overlap in discourse structure between the encoded entities lead to an encoding interference effect? Do linguistic entities sharing the same discourse structure information (e.g., both *the singers* and *the violinists* being part of the same main discourse structure) lead to an interference effect? Given the comparatively limited amount of work on the memory encoding process, and particularly the lack of discussion on the role of discourse information during encoding, it is an empirical question whether any interference effects will arise. This addresses the third research question.

In addressing the three questions (and accordingly conducting three sets of experiments), a self-paced reading task is used for the two sets of experiments, and an eyetracking method (visual world paradigm) is used for the final set of experiments. All experiments have been conducted in a web-based setting.

1.3 Outline of the dissertation

The dissertation begins with an overview of the theory currently widely adopted for understanding real-time language comprehension, namely the CONTENT-ADDRESSABLE CUE-BASED RETRIEVAL THEORY (**Chapter 2**). An overview of the core concepts and empirical evidence that supports the theory are presented. I then discuss the type of FEATURES that are employed in the theory and raise the question of whether these features can be expressed in a binary, static way.

Chapter 3 provides background on previous discussions on understanding discourse. Among different approaches to formalizing and representing discourse, primarily the perspective that views discourse as a structured representation of discourse units and their interrelations is adopted in this dissertation. The following section delves into the distinction between primary/main and secondary/subordinate discourse structure information. Subsequently, the motivation to use restrictive relative clauses (RRCs) and appositive relative clauses (ARCs) as a test case is provided. Finally, I offer an overview of existing research on discourse processing.

Chapters 4–6 present 8 main experimental findings. In **Chapter 4**, the first research question relating to retrieval is addressed. Preceding the experimental section, an overview of existing research concerning the processing of appositives and predictions regarding the experimental results is provided. In the main experiments, I leverage the NUMBER AGREEMENT ATTRACTION EFFECT while keeping the primary contrast of RRCs and ARCs. The RRC serves as the baseline given the robust attraction effect in the RRC condition reported in the literature. Novel findings regarding the presence or absence of the agreement attraction effect with the embedded ARC structure are presented. Results from the three self-paced reading task experiments demonstrate the consistent presence of the attraction effect in all three experiments with the RRC structure, replicating previous findings. However, the effect with the ARC structure is observed only in Experiments 2–3, not Experiment 1. Based on

these findings, I argue that identifying *discourse questions* and representing the incremental language processing through *active state of discourse questions* is a more appropriate approach than maintaining a binary, static way of representing linguistic features.

Chapter 5 addresses the second question, further exploring the use of discourse structure information during the retrieval process. While maintaining the key contrast of RRC vs. ARC, I make use of a sentential configuration that involves an intra-sentential PRONOUN RESOLUTION. The pronoun has two grammatically possible antecedents within the sentence, but the discourse status of the antecedents differs depending on the RRC vs. ARC structure. Prior to the main experiment, a summary of the key features of the tested construction is provided. This is followed by three self-paced reading task experiments. The findings show that discourse structure information affects resolving pronouns, even when it does not decide the grammaticality of constraining the choice of antecedent. Crucially, the results suggest that retrieving antecedents that are part of the same discourse structure information leads to greater processing costs than when they are part of distinct discourse structures.

Chapter 6 delves into the way discourse structure information affects the encoding process, addressing the third question. While a similar construction used in Chapter 5 is used, web-based visual world paradigm eyetracking experiments are conducted. Prior to the main experiments, an overview of a visual world paradigm is provided, including related work, and a comparison between a lab-based and web-based setting using the paradigm. It is then followed by a section that discusses the issue of COMPETITION (INTERFERENCE) EFFECT during memory encoding, and the time windows of interest where the effect would arise. The results suggest that the overlap in discourse-structure information between the two antecedents (i.e., RRC condition) leads to an interference effect, which is identified by smaller eye-gaze differences between the two target antecedents compared to the condition where there is no overlap (i.e., ARC condition). Additionally, the findings from the two experiments collectively show that this competition effect is manifested at a fairly early

stage when the second antecedent is encoded, even before retrieval is triggered.

Chapter 7 summarizes the experimental results and contains a general discussion of the findings. The three questions are revisited, and proposals and implications as well as limitations are discussed. I then conclude.

Relevant experimental material can be found here: https://osf.io/mn47y/.

CHAPTER 2

CUE-BASED RETRIEVAL MECHANISM IN SENTENCE COMPREHENSION

In understanding how humans parse sentences in real time, and what leads to cognitive load in comprehending certain structures than others, different theories have argued for different reasons for the cognitive basis that supports language comprehension mechanism.

To name a few of the widely discussed theories in the sentence processing literature, some have highlighted the role of learning and usage experience in producing and comprehending language. In this line of approach, the statistical knowledge of language is crucial in understanding the way humans comprehend language. Statistical knowledge and prior experience on co-occurrences of words (as in *lexical probabilistic model* (e.g., McDonald & Shillcock, 2003a, 2003b)) or syntactic structures (as in *structured probabilistic model* (e.g., Christiansen & Chater, 1994, 1999; Hale, 2001; Levy, 2008)) are key to what shapes human parsing. Along this account, the ease/difficulty in comprehension or parsing can be explained by the degree of SURPRISAL (entropy) or activation of the upcoming entity, which is accountable by statistical knowledge about language. Others find the cognitive basis of parsing in the memory architecture. Incoming linguistic information is STORED, MAINTAINED, and eventually INTEGRATED to parse sentences and resolve linguistic dependencies (e.g., Caplan & Waters, 1999; Gibson, 1998; Just & Carpenter, 1992). Given the limited capacity of working memory, more storage and integration costs are predicted to lead to parsing difficulty. Such parsing difficulty is commonly signaled by longer reading times in reading task experiments.

The cognitive mechanism I adopt in this work takes a memory-based approach, particularly the CONTENT-ADDRESSABLE CUE-BASED RETRIEVAL theory (R. L. Lewis & Vasishth, 2005; McElree, 2000; Van Dyke & Lewis, 2003). Under this framework, the parser is engaged in the process of ENCODING and MAINTENANCE of linguistic information, and finally RETRIEVAL of the relevant information. Linguistic information is encoded in memory in a

bundle of features (e.g., [+singular, −animate]) that can represent the incoming linguistic input. The encoded linguistic representation is maintained in memory—the parser needs to hold onto the stored information—until the relevant information needs to be retrieved. Once the parser reaches the retrieval site (e.g., verb position for resolving a subject-verb dependency), retrieval happens through a content-addressable fashion (McElree, 2000; McElree, Foraker, & Dyer, 2003), directly accessing the retrieval target through retrieval cues (e.g., {+singular, −animate}).

In the following sections, I lay out an overview of the cue-based retrieval theory (Section 2.1) and empirical findings that support the theory (Section 2.2). In Section 2.3, I discuss how linguistic representations are expressed in *features*. I raise the issue that representing features in a binary, static term has caveats, which calls for an alternative approach.

2.1 The architecture

Earlier studies have shown the limited capacity in working memory and its relation to the complexity in sentence comprehension (e.g., Caplan & Waters, 1999; Gibson, 1998; Just & Carpenter, 1992; Warren & Gibson, 2002). Given the limited amount or memory resources, how are humans capable of retrieving prior information when it needs to be retrieved?

One thought is that access to previous linguistic representation happens serially (e.g., Sternberg, 1966). This would predict that more time would be required to access prior representation when there are more memory units that need to be stored and maintained. In (1), for example, as the parser reaches the verb, *laughed*, the previously encoded noun, *the editor*, needs to be accessed to be integrated with the verb.

(1) A subset of conditions in Experiment 2 in McElree et al. (2003)

a. Control: No embedded structure

The editor laughed.

b. Embedded ORC

   The editor [that the book amused] laughed.

c. Embedded ORC with a PP

   The editor of the prestigious journal [that the book amused] laughed.

d. Embedded ORC & SRC

   The editor [that the book [that won the award] amused] laughed.

From a serial access approach, the speed of retrieving and accessing the NP at the verb would increase as the number of interpolated/embedded structures between the main subject-verb dependency increases (e.g., between *the editor* and *laughed*). The representational strength and distinctiveness are also expected to decrease incrementally.

However, McElree et al. showed that this serial access-based prediction was not borne out. Using a speed-accuracy trade-off (SAT) paradigm, they showed that while the representational strength got weaker (i.e., lower accuracy) with more interpolated materials before the retrieval, the speed to access prior linguistic material did not differ between sentences (1b–1d).[1] This non-varying retrieval speed but different accuracy across retrieval targets suggested that there is a capacity limit in the focus of attention within the working memory. Within the working memory span, information or linguistic representations that are less recent and are outside the capacity of focal attention are less privileged in terms of accessibility (McElree, 2006; Oberauer & Bialkova, 2009). Given the limit in the capacity to keep linguistic information under focal attention, the authors suggested that prior information is retrieved through direct access (rather than serial access) (R. L. Lewis & Vasishth, 2005). Resolving linguistic dependencies is a process of retrieving a target (e.g., the subject that corresponds to the verb in a subject-verb agreement) by directly accessing it (in a

1. A speed-accuracy trade-off (SAT) paradigm allows to estimate the strength of the representation that needs to be retrieved (accuracy, or *asymptote*), how long it takes for the information becomes initially available (speed, *intercept*), and how fast it takes from the initial access to the information to the retrieval of the full representation (*rate*).

content-addressable fashion) using retrieval cues.

Under this cue-based retrieval theory, when the parser forms and resolves the dependency in real time, it is engaged in the process of (i) encoding linguistic information, (ii) generating retrieval cues, and (iii) directly accessing the targeted information using retrieval cues.

Take the subject-verb long-distance dependency in (2) as an example. *The toy* and *arrived* form a dependency, and at the point the parser reaches *arrived*, the corresponding subject needs to be retrieved to resolve the dependency.

(2) Example from R. L. Lewis, Vasishth, and Van Dyke (2006)

Melissa knew that **the toy** from her uncle in Bogotá **arrived** today.

In doing so, first, the parser is involved in the encoding process. Processing *the toy* involves ENCODING of the incoming input, which can be expressed in feature bundles (3).

(3) Feature bundles (adopted from R. L. Lewis et al. (2006), Figure 1)[2]

a. Encoding of the representation of *the toy*

$$\begin{bmatrix} \text{CATEGORY:} & \text{NP} \\ \text{HEAD:} & \textit{toy} \\ \text{CASE:} & \text{nominative} \\ \text{NUMBER:} & \text{singular} \end{bmatrix}$$

b. Encoding of the expected predicate at *the toy*

$$\begin{bmatrix} \text{CATEGORY:} & \text{S} \\ \text{HEAD:} & \text{open} \\ \text{NUMBER:} & \text{singular} \\ \text{COMPLEMENT:} & \text{open} \\ & \vdots \end{bmatrix}$$

2. "Open" means that the head of the upcoming predicate is empty.

c. Cue generation at the retrieval site (*arrived*)

$$\begin{bmatrix} \text{CATEGORY:} & \text{S} \\ \text{HEAD:} & \text{open} \\ \text{NUMBER:} & \text{singular} \end{bmatrix}$$

Both the encoding of the NP representation (3a) and the expectation of the upcoming predicate (3b) are involved during this process. As the parser moves on to the next linguistic items (i.e., *from her uncle in Bogotá*), the encoded expectation-representation (3b) still remains in working memory. But, if, as McElree (2001) demonstrates, only around 1 item could receive the focus of attention, the encoded item now becomes outside the focus of attention. After this MAINTENANCE of prior representation, the parser reaches the RETRIEVAL site (i.e., *arrived*). At this point, the parser is involved in a cue-generation process. Retrieval cues are generated based on the given context and grammatical knowledge and the given related linguistic information, and these retrieval cues are a subset of the features that were used for encoding prior linguistic information. During this retrieval process, multiple candidates with features that share retrieval cues are accessed. In (3c), a subset of (3b), is used as retrieval cues, and *the toy*, the target of retrieval, is retrieved and integrated with the verb, *arrived*.

In the following section, I illustrate empirical cases where the cue-based retrieval theory has been adopted to account for different linguistic dependencies. They serve as evidence for the framework, particularly exhibiting a phenomenon referred to as the interference effect.

2.2 Evidence: interference effect

The content-addressable mechanism has been supported by empirical cases of similarity-based interference effects (e.g., McElree et al., 2003; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006, 2011). If the retrieval mechanism is content-addressable and cue-based, previously stored items can be activated when they share their feature specification with

retrieval cues. This leads to CUE-OVERLOAD (Nairne, 2002; Watkins & Watkins, 1975), where retrieval cues (partially) match with features of multiple items, even when these items are not necessarily retrieval targets; as a result, RETRIEVAL INTERFERENCE EFFECT is commonly observed as a consequence.

A well-known case of interference effect comes from empirical findings that are now commonly referred to as the NUMBER AGREEMENT ATTRACTION EFFECT (e.g., Wagers et al., 2009). Compare the sentences in (4). Both (4a) and (4b) are ungrammatical. However, compared to the latter, the former has been reported to be rated higher and read faster in a reading task around the verb region (e.g., Dillon, Mishler, Sloggett, & Phillips, 2013; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Pearlmutter, Garnsey, & Bock, 1999; Wagers et al., 2009). While ungrammatical sentences usually lead to longer reading times, that grammaticality penalty is reduced, illustrating a "grammatical illusion" (Phillips et al., 2011).

(4) Experimental material in Pearlmutter et al. (1999)[3]

a. Feature-mismatching distractor

*__The key__$_{[+\text{SG, }+\text{SUBJ}]}$ to *the cabinet*$_{[+\text{SG}]}$ __were__$_{\{\text{PL, SUBJ}\}}$ rusty from many years of disuse.

b. Feature-matching distractor

*__The key__$_{[+\text{SG, }+\text{SUBJ}]}$ to *the cabinets*$_{[+\text{PL}]}$ __were__$_{\{\text{PL, SUBJ}\}}$ rusty from many years of disuse.

The number agreement attraction effect has been taken as evidence to support the cue-based retrieval mechanism in comprehension. At the retrieval site (*were*), retrieval cues are generated, and prior items that match these retrieval cues in their features are activated. While *the key* could be the target for retrieval given its grammatical position, it is not the correct target since it does not match the retrieval cue in the number feature; neither the

3. The subscripts denote the linguistic features: SG (singular), SUBJ (subject), and PL (plural).

feature representation of *the cabinet* in (4a) matches the retrieval cue. However, in (4b), *the cabinets* partially matches the retrieval cue. While both *the key* and *the cabinets* do not fully match the retrieval cue, the distractor noun partially matches it, and this results in faster reading times in (4b) compared to (4a).[4] Given the nature of the retrieval process that it is content-addressable and cue-based, items in memory can be activated and available for retrieval even when they do not match all the retrieval cues. As we saw in the case above, this leads to an effect where the processing cost is mitigated by the presence of a partially feature-matching item in memory. This effect is commonly referred to as FACILITATORY INTERFERENCE EFFECT.

The interference effect can be seen even when there are multiple linguistic items that (partially) match the retrieval cues. In a long-distance dependency in (5), the corresponding subject needs to be retrieved at the verb, *was*. The retrieval cues are generated at the retrieval site: noun (N), singular (SG), and nominative (NOM). The grammatical retrieval target is *the student*; it has all the matching features that are used for retrieval cues. But *the exam*, even though it is not grammatically correct, also has features that match the retrieval cues, either to a greater degree (5a) or less (5b). A *distractor* item such as *the exam* leads to what is known as a FAN EFFECT, or INHIBITORY INTERFERENCE EFFECT, such that it leads to greater processing cost.

(5) Experimental material in Van Dyke and Lewis (2003)

a. Higher feature overlap

The secretary forgot that **the student**$_{[+\text{N},+\text{SG},+\text{NOM}]}$ who thought that *the exam*$_{[+\text{N},+\text{SG},+\text{NOM}]}$ was important **was**$_{\{\text{N,SG,NOM}\}}$ standing in the hallway.

4. Different modeling work has suggested different reasons that underlie the faster reading time in facilitatory interference effect. An activation-based account suggests that the effect can be understood in terms of race theory, where the overall reading times become faster in the competitive situation with multiple partially matching items in memory (Vasishth, Nicenboim, Engelmann, & Burchert, 2019). But see Nicenboim and Vasishth (2018) for an alternative approach with a different computational model.

b. Lower feature overlap

The secretary forgot that **the student**$_{[+\text{N},+\text{SG},+\text{NOM}]}$ who was waiting for *the exam*$_{[+\text{N},+\text{SG},+\text{ACC}]}$ **was**$_{\{\text{N,SG,NOM}\}}$ standing in the hallway.

In the current example, the high feature-overlap with retrieval cue with the feature of the retrieval candidate in (5a) has longer reading times at the retrieval site compared to the low feature-overlap condition (5b).

Interference effect can also arise during the encoding process when features of the item representations are shared across different items during encoding (and maintenance)—this is known as the ENCODING INTERFERENCE EFFECT. When similar features are used for representing elements in memory during encoding, the items may have weaker/degraded representations. A way to understand this interference effect is through the FEATURE OVERWRITING process. Under the assumption that features cannot be used exhaustively (Nairne, 1990), when a feature is used for multiple items, the feature in the preceding item gets overwritten. If item representations are construed as a vector with features with values (such as +1 or −1) assigned, feature overwriting brings the assigned values to 0. The changed value of the vector feature gives the previously encoded feature a weaker/degraded representation. Or, feature overwriting can have a more global impact on the representations in memory; in this case, it is not only the previously encoded item but other items held in working memory that can be impacted by feature overwriting (Oberauer & Kliegl, 2006).[5]

In sentence comprehension, the encoding interference effect is often exhibited as longer

5. Other forms of interference have been identified in the literature. For example, interference can happen when contextual/retrieval cues are not distinctive enough to select the target. This situation leads to confusion (or COMPETITIVE QUEING), and studies have found cases where the competing representations get activated as strongly as the target due to such confusion. Or, following the SUPERPOSITION theory, similar feature representation between items can lead to activation of connection weight between units, contributing to the representation of the prior item. In this theory, the interference effect can lead to processing benefits. Finally, FEATURE OVERWRITING, as explained above, is similar to superposition but in a different direction where feature-match leads to distortion/degraded representation of elements. Interference is greater when there is more feature overlap between items in feature overwriting. See Oberauer (2009) and Oberauer, Farrell, Jarrold, and Lewandowsky (2016) for an overview of different theories accounting for the interference effect.

read times (Gordon, Hendrick, & Johnson, 2001, 2004; Gordon, Hendrick, & Levine, 2002; Hofmeister & Vasishth, 2014; Rich & Wagers, 2020; Villata, Tabor, & Franck, 2018) or shorter eye-gaze towards the target (Sekerina, Campanelli, & Van Dyke, 2016).

For example, as shown in Gordon et al. (2001)'s study, the overlap in the NP features during encoding would lead to greater processing costs. All three sentences in (6) are the same except for the type of noun used as the embedded subject–either *the barber*, *you*, or *Ben*. The sentence in (6a) encodes subjects (i.e., *the banker* and *the barber*) that share the same type of NPs, i.e., both are "description"-type NPs. In contrast, the embedded subject (*you* or *Ben*) in (6b–6c) do not share the same type of NP with *the banker*.

(6) A subset of materials in Experiments 2 & 3 in Gordon et al. (2001)

a. Description

The banker that **the barber** praised climbed the mountain.

b. Pronoun

The banker that **you** praised climbed the mountain.

c. Proper name

The banker that **Ben** praised climbed the mountain.

Gordon et al. (2001) found that reading times were longer in (6a) > (6b), and in (6a) > (6b), at the *praised* and *climbed* regions.[6] Even when the type of NP is not used as a retrieval cue, the features of NP, specifically the overlap in those, can lead to an interference effect.

Further evidence shows that the encoding interference effect is realized even when the features used for encoding are not necessarily retrieval cues. Rich and Wagers (2020), for example, compared cases where the encoded NPs semantically overlap in different degrees. In the high semantic overlap (7a) condition, *the knife* and *the sword* have similar semantic

6. In their original study, (6a) and (6b) were compared in Experiment 2; (6a) vs. (6c) were compared in Experiment 3. Subject-extracted relative clause-version of the same set of conditions was also included to compare the processing cost between subject-extracted relative clause and object-extracted relative clause.

features; the feature overlap would lead to a degraded representation of these items during the encoding stage. In contrast, there will be less weakening of representations in (7b), where there is a lower degree of semantic overlap—*the knife* and *the shirt*.

(7) Experimental material in Rich and Wagers (2020)

a. High similarity condition

The knife that **the sword** was placed near had been recently sharpened.

b. Low similarity condition

The knife that **the shirt** was placed near had been recently sharpened.

Rich and Wagers found that the *high overlap* condition led to longer reading times, with the competition effect appearing as early as around *had been recently*, even before relevant semantic information needed to be integrated (*sharpened*). Hence, the encoding interference effect can be realized even before the retrieval site (cf. Villata et al., 2018).

2.3 Features: binary and static?

As illustrated in the previous section, different types of information can be used as features for encoding, and accordingly as cues for retrieval. In sentence processing studies, scholars have identified these features based on the notion of features used in linguistics that are useful for understanding linguistic dependencies (see Parker, Shvartsman, and Van Dyke (2017) for a discussion; cf. Martin and McElree (2009)).

To name some linguistic features identified in sentence processing studies, structural features (Arnett & Wagers, 2017; R. L. Lewis & Vasishth, 2005; Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011), morphosyntactic features including number (Dillon et al., 2013; Pearlmutter et al., 1999; Wagers et al., 2009), gender (Badecker & Straub, 2002; Cunnings & Felser, 2013; Cunnings & Sturt, 2014; Dillon et al., 2013; Lago

et al., 2015; Patil, Vasishth, & Lewis, 2016; Slioussar & Malko, 2016; Sturt, 2003; Villata et al., 2018), and case (Fedorenko, Babyonyshev, & Gibson, 2004; Logačev & Vasishth, 2012), and structurally configurational information (Franck, Lassi, Frauenfelder, & Rizzi, 2006; Kush, 2013; Kush, Lidz, & Phillips, 2015; Obata, Lewis, Epstein, Bartek, & Boland, 2010; Xiang, Wang, & Cui, 2015) have been reported to lead to an interference effect. Lexical semantic information such as animacy (Chen, Jäger, & Vasishth, 2012; Jäger, Benz, Roeser, Dillon, & Vasishth, 2015; Kwon, Ong, Chen, & Zhang, 2019; Van Dyke, 2007) and thematic or semantic plausibility (Cunnings & Sturt, 2018; Van Dyke, 2007; Van Dyke & McElree, 2006), and the type of referential form (Gordon et al., 2001, 2002) have also been known to lead to an interference effect. Negative polarity items have been engaged in leading to interference effects (Drenhaus, Frisch, & Saddy, 2005; Vasishth, Brüssow, Lewis, & Drenhaus, 2008; Xiang, Dillon, & Phillips, 2009).

In most cases, these features have been represented in a binary (or privative) term. Items can be represented with morphosyntactic features as [±singular], [±masculine], or [+subject]; or lexico-semantic information as [±animate], or even more specific to the encoded items as [±shatterable] (e.g., *a plate* is [+shatterable] while *a letter* is not). These features are static in that they are determined at any point of the incremental processing of the sentence.

However, the binary, static approach to features may face a challenge when linguistic relational information is considered (see Kush (2013) and Alcocer and Phillips (2012) for discussions). Consider the anaphoric dependency in (8), where the grammatical antecedent of the pronoun is constrained by certain constraints. Representing the possible antecedent as [+singular, +masculine], for instance, would be insufficient/incorrect to resolve the pronoun.

(8) Examples in Kush (2013)

a. John$_i$ likes him$_{*i/j}$.

b. John$_i$ likes himself$_{i,*j}$.

A way to address such issues is to keep the feature-based system but to motivate to generate features that reflect the structural relation between linguistic items that constrain anaphora. In Kush (2013), this is done so by postulating a [LOCAL] feature and specifying it with integers such as [LOCAL: 0] or [LOCAL: 1]. This approach moves away from a binary expression but keeps the feature-based approach to cue-based retrieval and expresses the relational information of the scope of retrieval by adding sub-feature information. However, the assignment of the integer is tightly associated with the syntactic hierarchy, which can be seen after the termination of the sentence. This means this approach can run into a *look-ahead* problem, where the parser needs to know the integer for the feature to be assigned while being involved in the incremental comprehension process. This is counter to what happens in incremental processing.

An alternative way is to keep the feature-based approach but impose a constraint on the type of linguistic items in the way they bear the feature. For example, in (9a–9b), the referential phrase (containing *the janitor*) is accessible as an antecedent for the pronoun (*he*) regardless of whether the phrase c-commands the pronoun or not. However, in (9c–9d), there is a constraint on the relation between the quantificational noun phrase (QP) (*any janitor*) and the pronoun (*he*) that the co-reference is predicated on the c-commanding relation: the QP does not bind the pronoun in (9c) whereas it does in (9d).

(9) Experimental material in Kush, Lidz, and Phillips (2015)

a. Referential-*but*

Kathi didn't think **the janitor** liked performing his custodial duties, *but* **he** had to clean up messes left after prom anyway.

b. Referential-*when*

Kathi didn't think **the janitor** liked performing his custodial duties, *when* **he** had to clean up messes left after prom anyway.

c. Quantifier-*but*

Kathi didn't think **any janitor** liked performing his custodial duties, *but* **he** had to clean up messes left after prom anyway.

d. Quantifier-*when*

Kathi didn't think **any janitor** liked performing his custodial duties, *when* **he** had to clean up messes left after prom anyway.

Kush, Lidz, and Phillips suggest that this relational constraint can be explained by the retrieval process of using an [ACCESSIBLE] feature, specifically by either activating or de-activating the feature during sentence comprehension. Following their argument, the referential NPs always bear this feature, and they can be retrieved for an upcoming pronoun without constraints. On the contrary, as for QPs, while they also bear the same feature, the feature is deleted once the parser reaches the end of the QP scope. They propose that this tracking and deleting of the feature is done so by "an automatic, dynamic update procedure: whenever the parser shifts to a higher level of embedding from its previous position, it should retrieve all QPs at the last level and de-activate their [ACCESSIBLE] features (p. 36)."

The idea about dynamic update and activation of features during sentence comprehension is worthwhile to note. In this project, I expand on this intuition and offer an account of the way features no longer become "accessible." More precisely, I provide an account of the way items in memory no longer get targeted by memory retrieval. I argue that discourse questions (similar to Questions under Discussion) play an important role; specifically, the *active state* of discourse questions is used for constraining the features that are targeted for retrieval. Under this approach, (i) the parser is sensitive to what discourse questions are active at the moment of retrieval, and (ii) only items in memory that are associated with active discourse questions are targeted for retrieval. In the following Chapter 3, some background discourse structure representation and how discourse questions are constructed will be presented.

CHAPTER 3

REPRESENTING AND PROCESSING DISCOURSE

There are numerous interpretations and ways to understand discourse and types of discourse information. This includes information packaging such as given and new information (B. J. Birner & Ward, 1998; Kaiser & Trueswell, 2004), viewing discourse as strategies and moves to achieve conversational goals (Bruce & Farkas, 2007; Farkas & Bruce, 2010; D. Lewis, 1979; Roberts, 2004), coherence relation between discourse units (Hobbs, 1985; Kehler, 2002), or structured representation of discourse units and their relations to one another (Asher & Lascarides, 2003; Polanyi, 1988).

It has also been extensively studied how discourse is processed in real time, such as how givenness of information (Kaiser & Trueswell, 2004), or focused or non-focused status of discourse entities affect processing differently (J. E. Arnold, 1998; Colonna, Schimke, & Hemforth, 2015; Kaiser, 2011), or how contextual effect or topichood affects sentence processing (Altmann & Kamide, 1999; Clifton & Frazier, 2018). Some studies examined the way discourse structure affects sentence comprehension (Duff, Anand, Brasoveanu, & Rysling, 2023; Göbel, 2019). I focus on the role of discourse structure, especially the division between *main discourse* versus *subordinate discourse* structure. In the current study, I make the division of the main versus subordinate discourse structure information by making use of two different relative clause (RC) structures in English, namely appositive relative clauses (ARCs) and restrictive relative clauses (RRCs).

What follows in this Chapter is different ways of representing discourse (Section 3.1), linguistic aspects of RRCs and ARCs (Section 3.2), particularly their syntactic and semantic representation and their discourse status, and a summary of previous work on discourse processing (Section 3.3).

3.1 Representing discourse

In their seminar paper, Clark and Schaefer (1989) identify some shared assumptions across the proposed models of discourse in the literature. One is the assumption of the notion of *common ground* (e.g., Stalnaker, 1978). This refers to mutual knowledge and information among participants of the discourse. This is a type of information that is presupposed and taken for granted among discourse participants. The second assumption is that participants make certain discourse moves to contribute to the discourse. Clark and Schaefer refer to this as *accumulation*, where participants make an effort to update the common ground or even "destroy" (D. Lewis, 1979, p. 339) the existing shared knowledge. Hence, we can assume that discourse essentially involves participants' engagement in modifying the shared knowledge in some way or another.

There are numerous approaches and theories to model discourse. While they are not mutually exclusive to one another, they vary in their emphasis on different aspects of discourse.

Some discourse models highlight the contrast in the informativeness in discourse, namely given vs. new. In these models, the goal of communication is for a speaker to deliver their way of perception of the world or knowledge to the hearer in a precise and effective way. To achieve this goal, a cooperative speaker will provide guidance as much as possible to the hearer to help them construct a representation of the world the speaker has in mind. To do so, the speaker packages information such as by making use of existing information, i.e., *old* and *given* information, as a ground to introduce and transmit *new* information (Ariel, 1985; Prince, 1979; Webber, 1981). This leads to the distinction between old/given and new information in the discourse model.

Some other models focus on the process of resolving the gap and discrepancy between discourse participants (Bruce & Farkas, 2007; Farkas & Bruce, 2010; D. Lewis, 1979; Roberts, 2004). Discourse is viewed similarly to a negotiation process where the main task is to identify the (un)shared knowledge space and resolve the differences in the mental state under

a certain discourse agenda. Discourse participants (i.e., interlocutors) bring certain topics to the "table" and negotiate and accept or reject ideas about those topics. For example, asking a question is considered one of the crucial discourse moves, as it helps interlocutors identify the gap in the knowledge state. Their engagement and discourse move to achieve the conversational goal are most highlighted in this line of discourse models.

Finally, discourse can be viewed as a structured representation composed of discourse referents or units that are connected to one another in certain relationships (e.g., Heim, 1982; Hobbs, 1985; Kamp, 1981; Polanyi, 1988). The relational aspect of discourse referents is the key characteristic of this line of thought. From this approach, representing a discourse can mean identifying the atomic discourse units and the relations that unite the smallest discourse pieces. These relations can be generated by linguistic elements (Kamp, 1981; Kamp & Reyle, 1983) as in Discourse Representation Theory (DRT), *coherence* relation between discourse units (Hobbs, 1985; Kehler, 2002), or rhetorical relations between discourse units as in the Segmented Discourse Representation Theory (SDRT) (e.g., Asher & Lascarides, 2003).

3.1.1 Segmented Discourse Representation Theory

I zoom into Segmented Discourse Representation Theory (SDRT) (Asher & Lascarides, 2003).[1] SDRT views discourse structure as a hierarchically structured representation of multiple discourse units, identified by discourse relations among them. This key notion is as conceptualized in (1), where R stands for the relation between the discourse units it takes.

(1) $R(\pi_1, \pi_2)$

A new discourse component attaches to an existing one mostly in one of two ways:

1. SDRT is similar to DRT but differs in the way it captures discourse relations. Discourse relations are expressed through a set of rhetorical relations rather than linguistic expressions.

Coordination or *Subordination* (2a).

(2) a. Coordinating relation

$\pi_1 \rightarrow \pi_2$

b. Subordinating relation

π_2

$\downarrow$

π_2

In *Coordinating* relation, the new discourse unit is characterized by *Continuation*, and *Narration* of the previous discourse unit. This is usually known to push the discourse forward, as represented in a horizontal line in (2a). Once a new discourse unit is attached to the existing structure, the previous discourse unit cannot be accessed granted only one exceptional discourse relation: discourse subordination. By discourse subordination, the speaker can bring the content back even after the conversation has progressed. When a new discourse unit is in a *Subordinating* relation with the previous one, it does not progress the discourse to a new direction but rather serves as the *Explanation*, *Elaboration*, and *Background* of the previous discourse unit, as expressed in a vertical line in (2b).

To illustrate, a short dialogue that talks about John's evening is presented in (3). Each discourse unit is annotated with π_n.

(3) Lascarides and Asher (2008), p. 8

a. π_1. John had a great evening last night.

b. π_2. He had a great meal.

c. π_3. He ate salmon.

d. π_4. He devoured lots of cheese.

e. π_5. He won a dancing competition.

These discourse units are intertwined with one another in certain relations. For example, the last four discourse units are elaborations on the first discourse unit about John having a great evening last night. The pieces of discourse information that John ate salmon and that he devoured lots of cheese are narrations of a preceding discourse unit that John had a great meal. The discourse relations between discourse units form a discourse representation. See the original Figure 3 in Lascarides and Asher (2008) for an illustration of the diagram that depicts the hierarchical discourse relations.

This rhetorical relation among discourse units can be expressed in Segmented Discourse Representation Structure (SDRS), and for the above discourse (3), a well-formed SDRS can be expressed as in (4). Note that (4) is slightly modified from the original illustration in Asher and Lascarides (2003) (Example (17) in p. 139). The final component, $LAST$, which denotes the final label that was added to the logical form, is omitted below since it can be removed when there is no confusion about signaling the last discourse component. Here, A is a set of labels ($A \subseteq \{\pi_1, \pi_2, ...\pi_n\}$); $\mathcal{F}$ is a function which assigns each member of A a member of Φ, which is the set of well-formed SDRS-formulae; K_{π_i} is the discourse representation structure of the corresponding discourse unit (π_i).[2]

(4) $< A, \mathcal{F} >$, where:

- $A = \{\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7\}$
- $\mathcal{F}(\pi_1) = K_{\pi_1}$
- $\mathcal{F}(\pi_2) = K_{\pi_2}$
- $\mathcal{F}(\pi_3) = K_{\pi_3}$
- $\mathcal{F}(\pi_4) = K_{\pi_4}$
- $\mathcal{F}(\pi_5) = K_{\pi_5}$
- $\mathcal{F}(\pi_0) = Elaboration(\pi_1, \pi_6)$

2. See Definitions 10 and 11 in Asher and Lascarides (2003) (p. 138) for a precise definition.

- $\mathcal{F}(\pi_6) = Narration(\pi_2, \pi_5) \wedge Elaboration(\pi_2, \pi_7)$
- $\mathcal{F}(\pi_7) = Narration(\pi_3, \pi_4)$

This SDRS can be represented in a Discourse Representation Theory-style of notation, and it can be found in the original Figure 4.3 in Asher and Lascarides (2003).

Question-based discourse representation

A way of representing the discourse structure is to view it as a set of (sub)question and answer pairs (Büring, 2003; Hunter & Asher, 2016; Riester, 2019; Roberts, 2012; Velleman & Beaver, 2015). From this view, constructing discourse (or adding a new discourse unit) can be construed as providing an answer to a question such as Question under Discussion (QUD) that threads the discourse. For example, the discourse units, "He had a great meal" or "He won a dancing competition" could be an answer to a question that asks, "What did John do on that lovely evening?" The question (Q) and answer (A) are connected in pairs, where they form a hierarchical relation. If the pairs are represented as a tree diagram, the pair that shares the same question/topic is labeled with the same subscript number, and the superordinating or subordinating relation is denoted in levels in numbering (e.g., 0.1.1 is subordinate to 0.1). An illustration of this can be found in the original Figure 3 in Riester (2019).

In the current study, I adopt this question-based discourse structure representation, where discourse is represented as a hierarchical relation of question-answer pairs between discourse units connected through certain discourse relations. These DISCOURSE QUESTIONS are similar to QUDs such that they address the issue/topic that is most highlighted. However, discourse questions can be associated with both primary and secondary discourse information. For example, content inside an appositive relative clause is secondary, side-commentary information from an overall discourse perspective, but it adds new information as the discourse

is constructed. From an incremental discourse-construction perspective, this secondary information can serve as a meaningful answer to a discourse question. In this sense, discourse questions can be paired with and associated with secondary, side-commentary information. This approach can be used to represent the discourse status of linguistic components at a sentence level, e.g., sentences embedding RRCs and ARCs. This will be further elaborated in the following Section 3.2.

3.2 Restrictive and appositive relative clauses

As previously introduced, the current project focuses on the division of primary/main and secondary/subordinate information in discourse, where some discourse units are superordinate/subordinate to others. The hierarchical relationship between discourse units can be represented as sets of a question-answer pair.

The division in the degree of significance in discourse can be identified within a sentence as well, specifically in a structure that involves a restrictive relative clause (RRC) and an appositive relative clause (ARC). The structures embedding the key contrast are presented below in (5). The information that *the waitress* "sat near the girl" has a different degree of significance in discourse depending on how it is packaged. It is crucial, primary information that contributes to the main discourse when packaged in an RRC (5a), while it is secondary information that is part of subordinate discourse when positioned within an ARC (5b).

(5) a. The waitress *who sat near the girl* was unhappy. [RRC]
b. The waitress, *who sat near the girl*, was unhappy. [ARC]

In the following sections, I summarize previous arguments on the contrasting (and similar) properties between RRC and ARC. These will be discussed from a syntactic, semantic, and discourse perspective.

Separation of ARC from the main clause

The first line of thought suggests that ARCs are syntactically isolated from the main clause syntactically and represented at a distinct dimension semantically. Syntactically speaking, ARCs can be represented independently of the main/host clause (referred to as the *orphanage approach*). Semantically speaking, ARCs can be construed to contribute to a different semantic dimension than the main clause (known as the *two-dimensional* approach).

Supporting evidence for an orphanage approach comes from the observation made the difference in the ambiguous meaning in (6). Depending on the type of the relative clause, the ambiguity of *who* refers to changes. In (6a), what Sandy did is ambiguous as it could either mean Sandy recognized the same man who took Kim's wallet, or Sandy recognized the person who took Sandy's wallet (but not necessarily the one who took Kim's wallet). However, this ambiguity is no longer present in (6b). The content inside the ARC is not interpreted as part of the ellipsis, and "so did Sandy" would only mean that Sandy recognized the man. The absence of ambiguity (unlike in the RRC counterpart) is explainable if ARC is not available to resolve the ellipsis and holds a status as a separate sentence.

(6) D. Arnold (2007), p. 275

a. Kim recognized the man who took her wallet, and so did Sandy. [RRC]

b. Kim recognized the man, who took her wallet, and so did Sandy. [ARC]

This was one of the empirical observations that motivated the *orphanage approach* in syntactic theories. ARCs and main clauses are linked in terms of abstract sense or later only at the discourse level and do not have a shared structure (Canac-Marquis & Tremblay, 1998; Fabb, 1990; Safir, 1986). A similar but less radical version of this approach suggests that ARCs do have syntactic configurations and they are conjoined with the main clause by some type of coordination, but they are nonetheless disconnected from the main clause structure, and the relation between the ARC and the main clause remains independent (Emonds, 1979;

McCawley, 1982; Ross, 1967).

Similar ideas have been proposed from a semantic point of view, where ARCs and main clauses are represented at separate semantic dimensions (Bach, 1999; Potts, 2005; cf. Potts, 2012). This *two-dimensional* assumes that ARCs and main clauses contribute to a distinct semantic dimension in meaning. For instance, the two sentences in (7) have opposite meanings, whether Edna started the descent or not. However, despite the opposite meaning of the two sentences, the additional information that Edna is a fearless leader remains intact in both sentences. This is taken as evidence that the main clause content—Edna started/didn't start the descent—and the ARC content—Edna is a fearless leader—contributes to different semantic dimensions.

(7) Koev (2022), p. 5

a. Edna, a fearless leader, started the descent.

b. Edna, a fearless leader, did not start the descent.

ARCs are also known to be *not-at-issue*, as opposed to *at-issue* (Beaver, Roberts, Simons, & Tonhauser, 2017; Simons, Tonhauser, Beaver, & Roberts, 2010; Syrett & Koev, 2015), in that it is not the main point of the utterance. Consider the short dialogue between two speakers, presented in (8)—The hashtag (#) indicates infelicitousness. A direct rejection of an utterance such as by saying "That's not true" is felicitous when the utterance is related to a main assertion but less so targeting a not-at-issue content. Given the prior utterance in (8a), it is a natural continuation to respond to it with (8b) but not with (8c). The different naturalness of rejecting an at-issue versus not-at-issue content suggests an ARC content contributes to a separate meaning dimension apart from the main asserted utterance.

(8) a. Speaker A: Edna, a fearless leader, started the descent.

b. Speaker B_1: That's not true—Edna has not started the descent.

c. Speaker B_2: #That's not true—Edna is not a fearless leader.

Along a similar line, the class of appositives has also been argued to perform speech acts independently from the hosting matrix clause (Frazier, Dillon, & Clifton, 2018; Koev, 2022), or demonstrate distinct prosodic contours (Dehé & Kavalova, 2007; Truckenbrodt, 2015). Specifically, ARCs in English, as opposed to RRCs, are commonly marked with an intonational boundary before the RC, usually signaled with a pause (Dehé, 2014; Watson & Gibson, 2004).

Integration of ARC with the main clause

Some others view ARC to be more integrated with the main clause, or the main assertion. From a syntactic point of view, the proposal follows that the anchor (i.e., the antecedent to which the relative clause attaches) and the following relative clause form a single constituent or chunk for both ARCs and RRCs (de Vries, 2006; Griffiths & de Vries, 2013; Jackendoff, 1977). de Vries (2006) label this line of thought as the *constituency approach*. Both the ARC and the RRC content are part of the whole structure. However, ARCs and RRCs differ in terms of the position in the structural hierarchy that the relative clause attaches to the anchor, where RRC is adjoined to an NP while the ARC to a DP, is higher than the NP level. Evidence comes from sentences such as in (9), wherein (9a), the sentence implies that there is a group of lecturers that did not pass the test whereas in (9b) it implies no group of lecturers failed the test. ARC is to be positioned outside of the scope of the quantifier (*all*), adjoining at the DP level.

(9) Adapted from de Vries (2006), p. 234

a. All the lecturers who passed the test looked happy. [RRC]

b. All the lecturers, who passed the test, looked happy. [ARC]

Additionally, given other empirical findings, ARCs seem to contribute to the main assertion and the meaning of the entire sentence (Asher, 2000; Boër & Lycan, 1976). For instance, Syrett and Koev (2015) conducted human participant judgment experiments and showed that the truth value of the ARC indeed affects the truth value of the whole sentence. The authors presented four types of sentences in an ARC structure, varied by the truth value of the main clause and the ARC: true-true, true-false, false-true, and false-false (main clause-ARC). They found that the falsity of the ARC as well as the main clause content affects participants' judgment on the truth value of the whole sentence. Similarly, in the examples in (10), the ARC content in sentences (10a) and (10b) contribute to the meaning of the full sentence, truth-conditionally functioning the same as the sentence in (10c), as conjunction—*the conjunction approach*. This suggests that ARCs are not inherently separated from the main assertion.

(10) Boër and Lycan (1976), p.16

a. Dick, who is an expert on Austin, loves the Bonzo Dog Band.

b. Dick, who loves the Bonzo Dog Band, is an expert on Austin.

c. Dick is an expert on Austin and loves the Bonzo Dog Band.

Similar accounts (but different theoretical mechanistic details) have shown that the ARC content contributes to the shared information just as the main clause does (AnderBois, Brasoveanu, & Henderson, 2015; Koev, 2019; Murray, 2014; Schlenker, 2013). A case of *boundary-crossing phenomenon*, where interpretation of both the main clause and the ARC is required to figure out the meaning of the whole sentence, is a case in point (Amaral, Roberts, & Smith, 2007; AnderBois et al., 2015; Nouwen, 2007). In (11a), the meaning of ARC needs to be taken into consideration in order to interpret the meaning of the main clause—resolving the pronoun, *her*. In (11b) as well the ARC content should be taken into account in order to interpret the meaning of the whole sentence. The boundary between the

main clause content and the ARC content needs to be crossed to interpret the meaning of the full sentence. To account for this phenomenon, some argue for a *unidimensional* approach (as opposed to *two-dimensional*), where the interpretation of the linguistic content, including the ARC is viewed as an incremental updating process, and the ARC content is integrated into a single representation of the sentence and discourse along with the main clause content.

(11) AnderBois et al. (2015), p. 98

a. John, who had been kissed by Mary, kissed her too.

b. John kissed Mary, who kissed him too.

While the specific approach to understanding the syntactic representation of ARC structures is still under debate, a consensus in that ARCs are represented differently than RRCs, either syntactically separate or integrated at a syntactically higher level. ARCs contribute to the semantics of the main assertion differently than RRCs. However, they are not totally isolated from the main content as they contribute to the truth value of the sentential meaning and need to be interpreted to resolve linguistic dependencies, for instance.

3.2.1 Discourse status

In this section, I examine the distinction between ARCs and RRCs at a discourse level, viewing them from a perspective of discourse structure. I focus on (i) the subordinating relation between the ARC and the main discourse, and (ii) the different types of discourse questions that ARCs and RRCs are associated with.

Let us return to the key contrast, repeated in (12). An important feature of subordinate discourse units (e.g., "who sat near the girl" as in (12b)) is that they do not push the discourse forward in dialogue (Asher & Lascarides, 2003; Hunter & Asher, 2016; Jasinskaja, 2016; Riester, 2019).

(12) a. The waitress *who sat near the girl* was unhappy. [RRC]

b. The waitress, *who sat near the girl*, was unhappy. [ARC]

This is exemplified in (13), where speaker A makes an utterance and speaker B replies to it. Speaker B_1's reply is a more natural continuation of speaker A's utterance than speaker B_2's reply. According to the discourse structure framework (Asher & Lascarides, 2003), this is because only the main discourse unit has the potential to move the discourse forward (the *Right Frontier Constraint* (Polanyi, 1988)) but not the subordinate discourse unit, and there is a preference for relating the upcoming material to the main assertion of the previous utterance (Frazier & Clifton, 2005; Jasinskaja, 2016).

(13) a. Speaker A: [The waitress, [who sat near the girl]$_{\text{SUBORDINATE}}$ was unhappy.]$_{\text{MAIN}}$

b. Speaker B_1: "I wonder why she was unhappy."

c. Speaker B_2: ?"I wonder why she sat next to her."

Given this subordinating discourse relation of ARC to the main clause, I compare the discourse status and the associated discourse questions of ARCs and RRCs. In doing so, I adopt the structured representation of discourse (Segmented Discourse Represented Theory) (Asher, 1993; Asher & Lascarides, 2003; Lascarides & Asher, 2008) and question-based approach to discourse (e.g., Hunter & Asher, 2016; Jasinskaja, 2016; Riester, 2019), to represent the discourse relation between ARC and main discourse content. This approach follows three principles.

First, each discourse unit, whether main or subordinate, can introduce a new *discourse question*. The term, "discourse question," is used in a broader sense than QUDs, encompassing questions related to both main utterances and side-commentary content. These questions are represented in a question-answer stack, associated with each discourse unit. Secondly, these questions are constructed and updated incrementally, as new discourse units

are added to the discourse. Thirdly, the discourse questions have different ACTIVE STATES, either *active* or *inactive*, and can be even POPPED OFF from discourse. A discourse question remains in an *active state* until resolved; and after the resolution of a discourse question, it will be removed (i.e., popped off) from the question stack. The extent to which the main and subordinate discourse units can interact with each other is determined by whether the discourse questions hosted by them are still in an active state.

Figure 3.1 illustrates a schematization of the incremental addition and removal of discourse questions in a sentence involving an ARC structure. Questions inside grey boxes are discourse questions related to linguistic inputs. These questions are built incrementally from left to right. The opaque question box indicates the discourse question is no longer active.

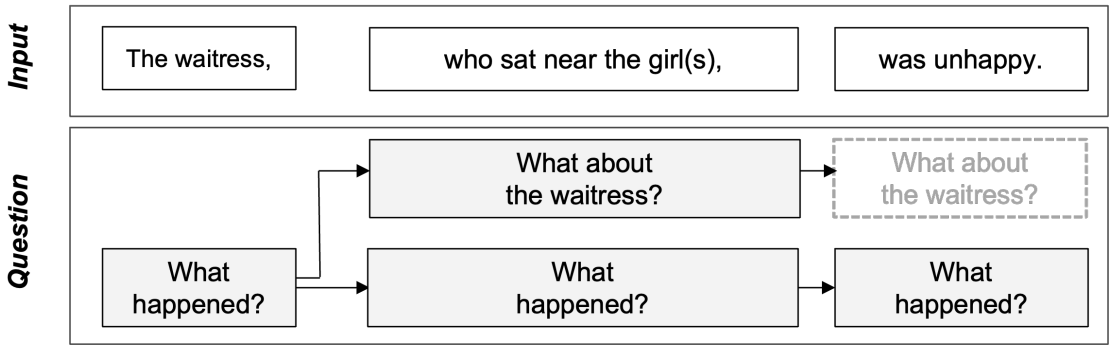


Figure 3.1: Discourse questions with an ARC structure.

The top panel shows the incoming information; the bottom panel illustrates the relevant discourse questions. When the subject noun phrase (*The waitress*) appears as the input, a discourse question about what happened (or "What happened to the waitress?") could be raised. This question stays active until it has been resolved. In the current sentence, this question will stay active throughout the sentence until the end, at which point the matrix clause (*the waitress was unhappy*) resolves the question. When the ARC (*who sat near the girl*) appears as the input, a new discourse question ("What about the waitress?") is added to the set of questions and stays active on the question stack until the ARC ends, at which point the question "What about the waitress?" is resolved by the information that the waitress

sat near the girl. This question is therefore deactivated and the question stack has only one active question left, i.e., the question associated with the main point of the utterance, "What happened?"

Regarding RRCs, treating them as intersective modifiers (Kratzer & Heim, 1998; Partee, 1975), it is assumed that they do not stand as independent discourse units but rather are subsumed under the discourse unit to which the RRC is attached. This is conceptualized in Figure 3.2. Questions inside grey boxes are discourse questions related to linguistic inputs, and the opaque box indicates the discourse question is no longer active.

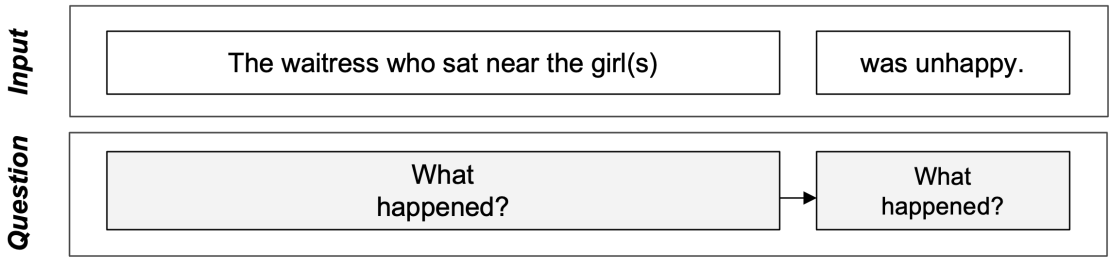


Figure 3.2: Discourse questions with an RRC structure.

As illustrated above, the RRC-associated discourse unit is part of the main discourse question ("What happened?") and does not give rise to a separate discourse question. Alternatively, it is also possible that RRCs can raise questions (e.g., "Which waitress?"), but due to their restrictive function, the question they raise would be a sub-question subsumed under the higher super-question ("What happened?"). This would be consistent with the experimental findings in Göbel (2019), where RRCs, in certain situations, can act similarly to subordinated discourse units. The super-question associated with the main discourse unit ("What happened?") remains active until it is resolved at the end of the sentence.

3.3 Discourse processing

There has been numerous work on how different aspects of discourse affect language use in real time. Some widely discussed factors in discourse include how givenness or accessibility of information (Ariel, 1985; J. E. Arnold, 1998, 2001, 2010; Gundel, Hedberg, & Zacharski, 1993; Kaiser & Trueswell, 2004; Lam & Hwang, 2022), or focused or non-focused status of discourse entities affect processing differently (Benatar & Clifton, 2014; S. Birch & Rayner, 1997; S. L. Birch & Garnsey, 1995; Blutner & Sommer, 1988; Colonna et al., 2015; Cutler & Fodor, 1979; Foraker & McElree, 2007; Kaiser, 2011; Ward & Sturt, 2007), or how topichood (including QUD) shapes discourse and affects language production and comprehension (Altmann & Kamide, 1999; Clifton & Frazier, 2018; Frazier & Clifton, 2005; Rohde & Kehler, 2014).[3]

Earlier work has also shown that discourse factors affect the degree of memory attainment of entities. S. L. Birch and Garnsey (1995) showed that entities that are focused have an advantage in enhanced memory. The authors used different sentential structures to insert the target word (italicized in (14)), either with a focus or non-focus structure. In the focus condition, the target word was positioned in either an *it*-cleft sentence or a there-insertion sentence whereas in the non-focus condition, it was placed in different syntactic positions.

(14) An example set of target sentences in S. L. Birch and Garnsey (1995)

a. Focus (*it*-cleft) and non-focus pair

(i) Focus condition: "It was the *singer* who attracted such large crowds to the nightclub."

(ii) Non-focus condition: "The donation from the *singer* would be used to buy food and medicine."

3. See J. E. Arnold (2010) for an overview of how *accessibility* can be interpreted in perspectives of *givenness*, *recency*, *syntactic prominence*, and *thematic prominence*.

b. Focus (*there*-insertion) and non-focus pair

(i) Focus condition: "There was this *hunter* who had tracked down a very dangerous animal."

(ii) Non-focus condition: "Jake is a *hunter* and avid fisherman who simply loves the outdoors."

Using a speeded (Experiment 1) and delayed (Experiment 2) word recognition task, the authors presented a stimuli sentence containing a target word and then showed the participants a test word (either identical, similar, or unrelated to the target word). The participants were asked to respond either "yes" or "no" as to whether they saw the test word in the previous sentence. The authors found that target words that were focused had faster response time and showed higher accuracy compared to those that were in a non-focused structure.

Using different experimental measures, previous work has also shown that topichood, for instance, is essential in understanding the selection and production of pronouns. Rohde and Kehler (2014) conducted production experiments using a written completion task and compared different influential factors that guide pronoun production. They specifically contrasted topichood and subjecthood, where the former in their definition refers to the syntactic object that takes the topical position and the latter is the sentential subject. Using the key contrast in (15), they compared the preference for the subject (*she*) referring to either Amanda or Brittany. In their experiment, two conditions without the pronoun prompt (*she* in the following sentence) were also included.

(15) An example story-continuation prompt in Rohde and Kehler (2014)

a. Amanda amazed Brittany. She _____

b. Brittany was amazed by Amanda. She _____

The authors particularly used subject-biased implicit causality verbs such as *amaze*, a family

of verbs that bias the discourse towards re-mentioning entities that provide the cause or the reason of the event. Making use of these verbs, the authors hypothesized that if pronoun production is driven by subjecthood, the discourse after the given prompt sentence will progress by referring *she* to the subject, i.e., Amanda in (15a) and Brittany in (15b). However, if topichood, i.e., topicality assigned by the syntactic subject status, is what drives pronoun production, preference for *she* to refer to Amanda in both (15a) and (15b) is predicted. The second hypothesis was borne out, with the results indicating that the entity with topichood was selected more than 60% of the time in both structures.

QUD is also known to guide language comprehension in resolving ambiguity. Kehler (2015) shows that QUD can determine the type of interpretation of a verb phrase ellipsis in English. The sentence in (16) can have two different interpretations, where one interprets as that Bill read John's paper (also known as the *strict* reading) whereas in the other case, Bill read his own paper (usually referred to as the *sloppy* reading).

(16) Two possible readings of verb phrase ellipsis (Kehler, 2015)

John read his paper, and Bill did too.

a. Who read John's paper?

John$_i$ read his$_i$ paper, and Bill$_j$ did ~~read John$_i$'s paper~~ too. [strict reading]

b. Who read his paper?

John$_i$ read his$_i$ paper, and Bill$_j$ did ~~read his$_j$ paper~~ too. [sloppy reading]

Depending on the QUD, however, the author suggests that this ambiguous reading can be resolved. In the case where the QUD pertains to addressing a list of candidates that read John's paper (16a), Bill as well as John becomes part of the alternative set for the answer. This confines the reading to be interpreted to mean that Bill read John's paper. In contrast, the question in (16b) gives rise to a set of answers where x read *x's paper*, and this leads to the reading of Bill read his own paper. Other studies present similar findings (e.g., Clifton

& Frazier, 2018; Frazier & Clifton, 2005), where even covert (as opposed to explicit) QUD also has a similar effect on resolving ambiguity (e.g., Clifton & Frazier, 2012, 2018; Grant, Clifton, & Frazier, 2012).

There are also studies that focus on the way discourse structure affects sentence comprehension (Duff et al., 2023; Göbel, 2019). Göbel (2019) specifically examined the way discourse relation affects the interpretation of pronominals. While using an ARC structure, the author varied the discourse relation between the ARC content and the main clause by using a past progressive (e.g., *was selling*) for the subordinating condition (17a) whereas a simple past (e.g., *sold*) and a temporal adverb (e.g., *then*) for the coordinating condition (17b). After being presented with either condition, the participants were asked to choose the antecedent of a pronoun, either *she* or *that*, that follows after the target sentence. There were two possible options, either *the real estate lady* (associated with the main clause) or *the landlady* (associated with the ARC content).

(17) A sample material in Göbel (2019) (Experiment 2)

a. Subordinating relation

At the open house, *the real estate lady* haggled with the potential tenants, who *the landlady* was selling some furniture to.

b. Coordinating relation

At the open house, *the real estate lady* haggled with the potential tenants, who *the landlady* then sold some furniture to.

The author found that in the subordinating discourse relation, the referent in the main clause (*the real estate lady*) was selected as the antecedent about 60% of the time; however, in the coordinating relation, this preference dropped to about 45%. The results showed that discourse structure and the relation between discourse units affect the interpretation of discourse, specifically anaphora resolution.

The current work will be an extension to the previous studies that examined different discourse factors that affect syntactic parsing and interpretation and resolution of pronouns. Similar to Göbel (2019), I manipulate the discourse status of linguistic entities by making use of the contrasting RCs in English. As illustrated in the previous section, RRCs will be used for establishing the main discourse relation (similar to coordination) whereas ARCs for subordinate discourse relation (as in the subordinating discourse relation). I examine the way these two distinct discourse structures and the relation between units affect syntactic dependency resolution, long-distance subject-verb dependency in particular, and pronoun resolution in real time.

CHAPTER 4

RETRIEVING DISCOURSE STRUCTURE INFORMATION: PART I

In this Chapter, I examine the validity of [+MAIN] and [+SUBORDINATE] features as retrieval cues in real-time comprehension. Doing so involves addressing two specific questions. One concerns the broader question of whether discourse structure-related information is used during the retrieval process, specifically, constraining the dependency-building process. If so, the other tackles the question of the way discourse structure information is represented and used in working memory: in a binary way—either [+FEATURE] or [−FEATURE], or in a dynamic mode where the discourse status of linguistic items is dependent on the active state of discourse questions (as was illustrated in Section 3.2.1).

To address these questions, I make use of a structure that requires retrieval of previous linguistic items for a successful parsing of the sentence. One of the most commonly used structures is the one that involves a long-distance dependency such as subject-verb number agreement with intervening linguistic materials. Additionally, to understand different linguistic features that are utilized during the retrieval process, I use a structure that is known to evoke a number agreement attraction effect. The attraction effect has been used as a hallmark case for demonstrating when and how linguistic features (e.g., gender, number, animacy) are employed as retrieval cues. Taking all these components together, I design a structure such that the distinction in discourse structure information is reflected while maintaining the key structure that involves a subject-verb agreement that induces a canonical number agreement attraction effect.

In Section 4.1, previous work on real-time comprehension of appositives is summarized. Empirical findings suggest that the processing cost with ARCs (or broadly appositives) is reduced compared to embedded RRCs; however, cases are reported where ARCs and RRCs do not differ in processing cost depending on the type of linguistic dependency and the

locus of retrieval. Section 4.2 provides an overview of the three sets of experiments and the predictions. Three self-paced reading task experiments are presented in Sections 4.3–4.5. Across three experiments, the discourse status (main vs. subordinate) of linguistic elements as well as the active status of the associated discourse questions are manipulated. The results collectively show that while discourse structure information is indeed utilized for dependency building, it is not the binary approach to discourse status but the active status of discourse questions that affects the incremental resolution process. This argument is spelled out more in detail in Section 4.6, wherein a comparison with different approaches is presented; limitations and future work are also discussed in this section. Section 4.7 concludes.

4.1 Previous work: processing of subordinate ARC structures

A number of studies have examined whether ARCs (or broadly appositives) are represented and processed differently compared to RRCs. Earlier work has shown findings that suggest independence of the ARC content from the main clause, taking evidence from reduced processing cost with embedded appositives compared to the control RRC structure (Dillon, Clifton, & Frazier, 2014; Kroll & Wagers, 2019). However, later investigations have found evidence that ARCs are not completely separated from the main content (Dillon, Clifton, Sloggett, & Frazier, 2017), with observations that ARCs may have similar linguistic representations as RRCs do (Dillon, Frazier, & Clifton, 2018). Using more time-sensitive measures, recent work has found that subordinate discourse information can affect parsing similarly to main discourse information when incremental processing is considered. These studies have shown that the independence of ARC can be affected by the timeline of processing (e.g., whether it is before or after the critical retrieval site) (Dillon et al., 2017) or the type of dependency (Ng & Husband, 2017).

Separation of main and subordinate discourse information

Some previous work has suggested that main and subordinate discourse information appears not to interact with each other during processing (Dillon et al., 2014; Kroll & Wagers, 2019).

Dillon et al. (2014), for example, first looked at the differences between the two conditions in (1). Having (1a) as a control condition, the authors examined the processing cost of the appositive counterpart (1b). Given the existing large body of work that longer embedded clauses incur greater processing cost (e.g., Gibson, 1998, 2000; Grodner & Gibson, 2005; Hale, 2001; Levy, 2008; R. L. Lewis & Vasishth, 2005; McElree et al., 2003; Van Dyke & Lewis, 2003; Warren & Gibson, 2002), the authors questioned whether the processing cost will be reduced in the presence of intervening subordinate discourse information as in (1b).

(1) A subset of the experimental material in Experiment 1 in Dillon et al. (2014)

   a. That butcher **who was in the busy shop (Amy visited on Third Avenue)** bought his meat from local farmers. [RRC] (Long-condition in parenthesis)

   b. That butcher, **the one in the busy shop (Amy visited on Third Avenue)**, bought his meat from local farmers. [Appositive] (Long-condition in parenthesis)

In their acceptability rating task, Dillon et al. (2014) found a length penalty effect in the baseline RRC condition (1a), where the longer intervening RRC incurred more processing burden than the shorter intervening RRC structure. Critically, however, the observed length penalty effect with the RRC condition was reduced in the appositive condition. Based on this finding, the authors argued that appositives and main clauses may be "processed independently [...] in separate memory store[s]" (Dillon et al., 2014).

Similar results have been reported in Kroll and Wagers (2019) and Duff et al. (2023). Kroll and Wagers (2019), for example, found the same asymmetric length penalty effect between RRCs and ARCs. Short and long embedded RCs were both included, and ARCs were marked with parentheses as shown in (2). A chunk of words was presented in different

windows, marked with a pipe (|); the second window included the critical RC content.

(2) Experimental material in Experiment 6 in Kroll and Wagers (2019)

a. The bear | **that is standing on the ball [the trainer rolled across the room]** | is wearing a hat. [RRC] (Long-condition in square brackets)

b. The bear | **(who is standing on the ball [the trainer rolled across the room])** | is wearing a hat. [ARC] (Long-condition in square brackets)

While the 2-way interaction of clause and embedded RC-length did not reach significance, they found the same pattern as in Dillon et al. (2014) where ARCs are rated higher than RRCs, particularly in the long-embedded-RC condition. These findings have been taken as evidence to argue that processing ARC content is inherently different than processing the main clause.[1]

Interaction of main and subordinate discourse information

However, some later work, while replicating the findings that information contained in main and subordinate discourse units do not interact with each other, also found evidence that main and subordinate discourse information is not entirely separated.

Consider the example in (3), tested in Dillon et al. (2017). Their experimental items had a *wh*-filler, who_1, which was connected to a sentence-final gap, *dinner for* ___. Additionally, there was an intervening *wh*-dependency (who_2 *bought Italian ham*) which was varied by the type of RC structure, either an RRC (3a) or an ARC (3b).

(3) Experimental material in Experiment 1 in Dillon et al. (2017)

1. Kroll and Wagers compare different possibilities that might contribute to the disparate nature of the two types of clauses/discourse information. They argue for the argument that leans into the distinct prosodic structure of appositives. I turn to this argument later in Section 4.6.4 and suggest that this may not hold true given the current empirical findings.

a. The butcher asked who_1 the lady who_2 bought Italian ham was cooking *dinner for* ___. [RRC]

b. The butcher asked who_1 the lady, who_2 bought Italian ham, was cooking *dinner for* ___. [ARC]

The eye-tracking-while-reading experiments in Dillon et al. (2017) revealed two interesting findings. First, there was a processing cost at the gap position (*dinner for*) in the RRC condition relative to its less complex baseline control that did not involve a *wh*-filler-gap dependency (e.g., "The butcher asked *if* the lady(,) who ..."). This was an expected syntactic complexity effect (Frazier & Clifton, 1989; Gibson & Warren, 2004; Gordon et al., 2001; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005; Staub, 2010; Wagers & Phillips, 2014). The corresponding complexity effect, however, did not arise in the ARC condition. The contrast between RRC and ARC at the sentence-final gap position replicated the general findings in Dillon et al. (2014). Meanwhile, the other important finding was that there was no difference between ARC and RRC conditions in the intervening RC regions. The intervening material (specifically on the "who$_2$ bought" region) triggered a reading time slowdown in both ARC and RRC conditions relative to their respective baseline controls, presumably because there was interference from an open outer *wh*-dependency.

Dillon et al. (2018) further showed that ARC content, or subordinate discourse information, is not inherently separated from the main clause or main discourse information. This time they examined the attachment preference given a structural ambiguity as in (4).

(4) A subset of the experimental material in Experiment 1 in Dillon et al. (2018)

a. The journalist interviewed **the daughter** of the colonel(,) who was on the balcony.

b. The journalist interviewed the daughter with **the colonel**(,) who was on the balcony.

Previous work has shown that there is a preference for high attachment of the RRC to *the daughter* (4a) whereas low-attachment to the PP (*the colonel*) (4b) (e.g., Grillo & Costa, 2014). The authors hypothesized that if ARC content is linguistically (syntactically, in particular) integrated with the main clause just as an RRC, a similar attachment preference should be expected. The syntactic attachment preferences in RRCs and ARCs were shown in a similar fashion, suggesting that it is hasty to conclude that the ARC content is separate and independent of the main clause.

Other factors affecting real-time comprehension of ARCs

Empirical findings so far seem to also suggest that the type of linguistic dependency and the experimental task affect the processing of ARC structures.

Ng and Husband (2017) examined the processing of appositives in two different dependencies, namely the subject-verb agreement and the negative polarity dependency. They made use of the number agreement attraction effect and negative polarity item (NPI) illusion effect to examine the retrieval process of appositive content (see Chapter 2 for a summary of these effects). While processing difference, i.e., retrieval interference effect, was observed between the appositive and the control RRC structure in the attraction effect, the difference was not found in examining the NPI illusion.

Sentences in (5) present a set of examples used in Ng and Husband (2017) for comparing the number agreement attraction effect between an embedded RRC and appositive structures. The standard number agreement attraction effect with the RRC structure (5a) predicts that in the ungrammatical condition (*were*), the presence of a number feature-matching distractor NP (*the rock singers*) leads to decreased reading times compared to the absence of it (*the rock singer*) at the retrieval site (*were* or the following region) (e.g., Wagers et al., 2009).

(5) Material used for testing the attraction effect in Ng and Husband (2017)

a. **The demo tape** that promoted *the rock singer(s)* {**was**/*__were__} very successful in America in the 1990s. [RRC]

b. **The demo tape** the one that promoted *the rock singer(s)* {**was**/*__were__} very successful in America in the 1990s. [Appositive]

The authors replicated the attraction effect in the RRC condition. The effect, however, was absent in the appositive condition (5b)—the same finding has been replicated in a later study in McInnerney and Atkinson (2020) as well.[2]

Next, sentences in (6) illustrate the key contrast for examining the NPI illusion in the two distinct structures. An NPI expression such as *ever* is grammatically licensed when it appears in the scope of a negative licensor such as *no* (6a). But an NPI illusion has been observed such that a licensor outside of the scope (e.g., *no critics*) leads to reduced reading times despite the ungrammaticality (e.g., Vasishth et al., 2008; Xiang et al., 2009).

(6) Material used for testing the NPI illusion in Ng and Husband (2017)

a. {*__The__/**No**} **authors** that *no/the critics* recommended have **ever** received acknowledgement for a best selling novel. [RRC]

b. {*__The__/**No**} **authors** the ones that *no/the critics* recommended have **ever** received acknowledgement for a best selling novel [Appositive]

Unlike the case of the number agreement attraction effect, no reliable differences in terms of the NPI illusion effect were found between the RRC and appositive conditions.

2. Yet, unlike their Experiment 1, McInnerney and Atkinson (2020) found the number agreement attraction effect in both appositive and RRC structures in subsequent experiments (Experiments 2 and 3), where (non-)speeded acceptability judgment task was involved. The authors speculated that the mixed results may be due to the task effect, suggesting that the sensitivity to the division of main and subordinate discourse may be pronounced at different time periods, either during or after processing the sentence. But the proposal has remained unclear.

As to this discrepancy, Ng and Husband suggested that NPI licensing ultimately evokes semantic and pragmatic licensing mechanisms instead of merely a syntactic process, and hence the linguistic information inside the main and the subordinate discourse structure can be "bridged in pragmatically principled ways." If so, the division may not be distinct, and as a consequence, the retrieval process may be similar between the embedded RRC and appositive structures. This is different from the number agreement attraction effect, which is dependent on resolving the syntactic dependency. The type of linguistic dependency may have contributed to the similar or different processing between main and discourse structure information.

Remaining puzzle

The empirical findings seem to be that the distinction between main and subordinate discourse information is pronounced after the subordinate discourse information has been processed. The stark division between the two types of discourse is captured after the sentence has been processed completely as in judgment tasks (e.g., Dillon et al., 2014; Kroll & Wagers, 2019), and when the retrieval of relevant information takes place after the termination of encoding the subordinate content (e.g., Dillon et al., 2017; McInnerney & Atkinson, 2020; Ng & Husband, 2017). Yet, the division is not made during the encoding of the subordinate discourse information (e.g., Dillon et al., 2017). Once the parser progresses beyond the right boundary of the subordinate discourse unit (e.g., beyond the RC), however, the information within the subordinate discourse unit becomes more separated from the main discourse unit.

To account for these empirical findings, Dillon et al. propose that the rapid structural decay of ARC content after processing it could be a contributing factor, an idea also argued in Duff et al. (2023). Duff et al. posit that content within the subordinate discourse structure after the RC boundary becomes "discounted," exerting "less influence on downstream parsing [...] than other material" (p. 2). Similarly, Dillon et al. argue that the syntactic form of the

ARC structure gets "lost relatively rapidly, minimizing its ability to interfere with subsequent processing" (p. 105)—*structural loss hypothesis*. And the loss of the syntactic form of an ARC in working memory leads to a reduced interference effect between the main and the ARC content. There is some evidence from earlier work lending support to the general possibility that the surface form of a sentence is short-lived (e.g., Lombardi & Potter, 1992; Potter & Lombardi, 1990). These studies have shown that when recalling a sentence, instead of recalling from verbatim memory, people regenerate the form of the sentence based on the conceptual representation of it.

However, there is no direct evidence that the syntactic form of an ARC structure rapidly decays. It is also unclear why appositives should undergo syntactic loss when the RRC counterparts do not undergo a similar forgetting process. A potential argument is that RRCs and appositives may exhibit differential susceptibility to the forgetting process due to their syntactic difference, but imposing such a constraint would largely rely on stipulation. Moreover, empirical findings on the decay of syntactic forms are varied, with studies indicating that these forms can persist even after extended delays (Gurevich, Johnson, & Goldberg, 2010; Kaschak, Kutta, & Schatschneider, 2011).

4.2 Overview of experiments and predictions

In the current work, I focus on the earlier observation that the interactability between main and subordinate discourse information is dependent on whether processing ARC is closed off or not. Returning to the initial question I raised, I examine the two competing approaches to discourse structure information. One is to assume a binary division of the discourse structure, where linguistic content is construed to be either part of the [+MAIN] or [+SUBORDINATE] discourse. The other is to account for the dynamically changing status of discourse questions associated with the linguistic content. I compare these two approaches by manipulating the discourse status and the active state of discourse questions of the linguistic content

across three experiments (Table 4.1). Following the design in Ng and Husband (2017) and McInnerney and Atkinson (2020), I make use of the number agreement attraction effect to examine the retrieval interference effect when processing subordinate discourse information is involved. I varied the relative position of the retrieval target NP and the distractor NP such that the two NPs are part of the same discourse structure in the RRC condition but distinct structures in the ARC condition. In the meantime, the locus of the retrieval site was varied so that the active state of discourse questions was also manipulated.

Table 4.1: Prediction on Experiments 1–3 factored by discourse status and active state of discourse questions

Exp.	Clause	Discourse status [Target, Distractor]	Active state of discourse questions [Target, Distractor]	Number agreement attraction
1	RRC	[+ MAIN, + MAIN]	[+ ACTIVE, + ACTIVE]	yes
	ARC	[+ **MAIN**, + **SUBORD.**]	[+ **ACTIVE**, −**ACTIVE**]	?
2 & 3	RRC	[+ MAIN, + MAIN]	[+ ACTIVE, + ACTIVE]	yes
	ARC	[+ **SUBORD.**, + **MAIN**]	[+ **ACTIVE**, + **ACTIVE**]	?

Note. Subord. = Subordinate

These manipulations are exemplified below. In Experiment 1, the target (*the waitress*) and the distractor (*the girls*) were either part of the same main discourse (7a) or separate discourse structures (7b). As the retrieval site (*were*) was after the ARC had closed off, the active state of the distractor in the ARC condition was inactive; it was active in the RRC condition since the RRC contributed to the main discourse active question. Having this RRC structure as a control condition (e.g., Parker & An, 2018), I examine whether the same attraction effect is found in the ARC counterpart as in (7b).

(7) A sample sentence in Experiment 1 (ungrammatical, plural distractor condition)

a. *__The waitress__ who sat near *the girls* surprisingly **were** unhappy. [RRC]

b. *__The waitress__, who sat near *the girls*, surprisingly **were** unhappy. [ARC]

While the same binary division of [+MAIN] and [+SUBORDINATE] is made with RRC vs. ARC contrast, the critical retrieval site is located inside the RC in Experiment 2 (8). This makes the discourse question associated with the content inside the RC active, regardless of ARC or RRC. The sentence in (8a) also represents a canonical structure of a number agreement attraction effect (e.g., Wagers et al., 2009). It is of empirical question whether the same pattern of number agreement attraction effect will be observed in (8) as in (7). If the same pattern arises, it would suggest that the distinction in the discourse status—whether the NP is part of the main or the subordinate discourse—matters over the real-time status of the discourse question. If not, we can conjecture that the dynamically changing active state of discourse questions is what matters at the point of retrieval.

(8) A sample sentence in Experiment 2 (ungrammatical, plural distractor condition)

a. **The musicians* who **the reviewer praise** so highly will win a Grammy. [RRC]

b. **The musicians*, who **the reviewer praise** so highly, will win a Grammy. [ARC]

Experiment 3, as illustrated in Table 4.1, reflects the same experimental contrast as in Experiment 2. It is possible that the open dependency after the matrix subject in Experiment 2 (*the musicians*) kept the distractor NP in the relative clause at a relatively high activation level in both structures. To tease apart this possibility, in Experiment 3, the RC is positioned after the core argument, i.e., object, so that there is no open dependency at the point when the RC content is being processed. We expect to see the same pattern as in Experiment 2 if the parsing is not solely driven by the activation and attention level but rather a principled factor such as discourse status or active state of discourse questions.

(9) A sample sentence in Experiment 3 (ungrammatical, plural distractor condition)

a. *Alicia met *the musicians* who **the reviewer praise** so highly.

b. *Alicia met *the musicians*, who **the reviewer praise** so highly.

A self-paced reading task is used for all three experiments, each of which is accompanied by a comprehension question after each trial that asks about the content of the sentence. Having the RRC structure as the control condition, a number agreement attraction effect is predicted in the RRC condition in all three experiments. With this baseline, the empirical question is whether we would see the same effect in the ARC condition in different experimental designs.

To preview the finding, I replicate the attraction effect in all three experiments in the RRC structure. More importantly, the effect is absent in the ARC condition in Experiment 1 but present in Experiments 2 and 3. The findings collectively suggest that discourse structure is used during the retrieval process, but more precisely, the active state of discourse questions associated with the linguistic content is what matters during the retrieval.

4.3 Experiment 1

4.3.1 Methods

Subjects

A totla of 120 native speakers of American English residing in the U.S. via the Prolific platform. Participation recruitment was conducted with IRB approval from the local institution. Two participants were excluded from the main analysis as they self-reported that their first language was not English, leaving us 118 participants for the analysis (mean age = 30.94; range: 18–50). The duration of the experiment was approximately 20 minutes, and participants were paid 3.50 USD in compensation.

Material and design

Material for Experiment 1 is presented in Table 4.2. The material consisted of 48 items in a fully crossed 2 x 2 x 2 design with grammaticality (grammatical vs. ungrammatical),

distractor number (singular vs. plural), and clause type (ARC vs. RRC) as factors. The material with RRCs was adapted from Experiment 1 in Parker and An (2018). All the target sentences had the structure of *NP1(,) who VERB Preposition NP2(,) ADVERB {was/were} ...* (e.g., "The waitress(,) who sat near the girl(s)(,) unsurprisingly was/were unhappy about all the noise."). NP1, the target subject, was always a singular noun (e.g., *The waitress*). The main auxiliary verb was varied by grammaticality: *was* (grammatical) or *were* (ungrammatical). The target subject was always modified with a subject-extracted RC, which contained the distractor noun, NP2. The distractor noun was either a singular noun or a plural noun (e.g., *the girl(s)*). The RC was either an RRC or an ARC, and the ARC conditions were marked with commas before and after the RC boundary. There was always an intervening adverb between the distractor noun and the main verb (e.g., *unsurprisingly*). The number of regions for the target trials ranged from 11 to 13. Additionally, 24 filler sentences—all grammatical sentences—were included.

Procedure

A Latin-square design was used to assign the 48 main trial items into eight lists. Each participant read a total of 72 sentences (including filler sentences), with half of the sentences being grammatical and the other half ungrammatical. The presentation of the trials was randomized for each participant.

The experiment was conducted on IbexFarm (http://spellout.net/ibexfarm), an online experiment platform. Participants read sentences in a self-paced phrase-by-phrase moving window paradigm. Each sentence was presented individually on each participant's screen. Words were initially masked by dashes, and each word or phrase appeared as the participant pressed the space bar. The sentence was presented in a non-cumulative fashion, where the previous word was masked again by dashes as the participant proceeded to the next region. Participants were instructed to read the sentences as naturally as possible at their regular

Table 4.2: An example of the experimental item for Experiment 1

Condition	Sentence
ARC-Sg-Gr	The waitress, / who / sat / near / *the girl*, / unsurprisingly / **was** / **unhappy** / about / all / the noise.
ARC-Sg-Ug	The waitress, / who / sat / near / *the girl*, / unsurprisingly / **were** / **unhappy** / about / all / the noise.
ARC-Pl-Gr	The waitress, / who / sat / near / *the girls*, / unsurprisingly / **was** / **unhappy** / about / all / the noise.
ARC-Pl-Ug	The waitress, / who / sat / near / *the girls*, / unsurprisingly / **were** / **unhappy** / about / all / the noise.
RRC-Sg-Gr	The waitress / who / sat / near / *the girl* / unsurprisingly / **was** / **unhappy** / about / all / the noise.
RRC-Sg-Ug	The waitress / who / sat / near / *the girl* / unsurprisingly / **were** / **unhappy** / about / all / the noise.
RRC-Pl-Gr	The waitress / who / sat / near / *the girls* / unsurprisingly / **was** / **unhappy** / about / all / the noise.
RRC-Pl-Ug	The waitress / who / sat / near / *the girls* / unsurprisingly / **were** / **unhappy** / about / all / the noise.

Note. ARC = Appositive relative clause. RRC = Restrictive relative clause. Sg = Singular distractor noun. Pl = Plural distractor noun. Gr = Grammatical. Ug = Ungrammatical. Distractor noun italicized. Regions of interest are bold-faced. The '/' sign indicates regions.

reading pace. A 'yes/no' comprehension task that asked about the content of the sentence appeared after each sentence (e.g., "Was the waitress unhappy about all the noise?"). Half of the questions had 'yes' correct responses and the other half 'no.' The participants were not given any feedback on their responses to the task. Participants had 10 practice trials before the main experiment.

Analysis

Prior to data analysis, results from from 10 participants were excluded, whose accuracy on the comprehension question task for both target and filler items was below 80%. For the reading time analysis for the self-paced reading task, trials with incorrect comprehension question responses were removed for the main analysis (7.47% of the data). Reading time data points were removed that were beyond 3SD of the mean by condition and by region

(additional 1.5% removal of the data). Regions of interest included (a) the main verb position (*was/were*) (region 7) that agreed with the target subject (*the waitress*) and (b) the next spillover region (*unhappy*) (region 8).

The log-transformed reading time data was analyzed by conducting linear mixed-effects regression models, using the `lmerTest` package (Bates, Mächler, Bolker, & Walker, 2015) implemented in R (R Core Team, 2022). Fixed effects in the model included grammaticality (grammatical vs. ungrammatical), distractor number (singular vs. plural), clause type (ARC vs. RRC), and their interactions. Since the main interest of comparison is the presence (or absence) of the standard number agreement attraction effect between the two clause types, a nested contrast was used to examine the agreement attraction effect in ARC and RRC, following the approach in Nicenboim, Schad, and Vasishth (2023).[3] Using the nested contrast coding, the critical 2-way interaction between distractor number and grammaticality was estimated as a nested effect for ARC and RRC separately. To take into consideration the spillover effect from regions prior to the given region analysis, the log-transformed reading time of the immediately preceding region was also included as a fixed effect. All models were initially fit with maximal random effects structure, with by-participant and by-item random intercepts and by-participant and by-item random slopes for all fixed-effect predictors (Barr, Levy, Scheepers, & Tily, 2013). Random effects structure was simplified when the models failed to converge. I used the `buildmer` package (Voeten, 2022) in R to find the maximal model. A fixed effect was considered significant when the absolute t-value associated with the effect exceeded 2 (Gelman & Hill, 2006).

3. Also see Vasishth (2021) for the advantages of using nested contrasts; and see Patil et al. (2016) for a similar example using nested contrast for studying the interference effect.

4.3.2 Results

Comprehension accuracy

The mean comprehension question accuracy for each condition is presented in Table 4.3. The comprehension accuracy for all conditions was generally close to ceiling.

Table 4.3: Mean comprehension question accuracy in Experiment 1

	Gr-Pl	Gr-Sg	Ug-Pl	Ug-Sg
ARC	0.935 (±0.003)	0.925 (±0.003)	0.910 (±0.003)	0.937 (±0.003)
RRC	0.927 (±0.003)	0.934 (±0.003)	0.913 (±0.003)	0.922 (±0.003)

Note. Standard errors of the grand mean are in parentheses. ARC = Appositive relative clause. RRC = Restrictive relative clause. Sg = Singular distractor noun. Pl = Plural distractor noun. Gr = Grammatical. Ug = Ungrammatical.

Reading times

Figure 4.1 presents the mean log reading times for each region.

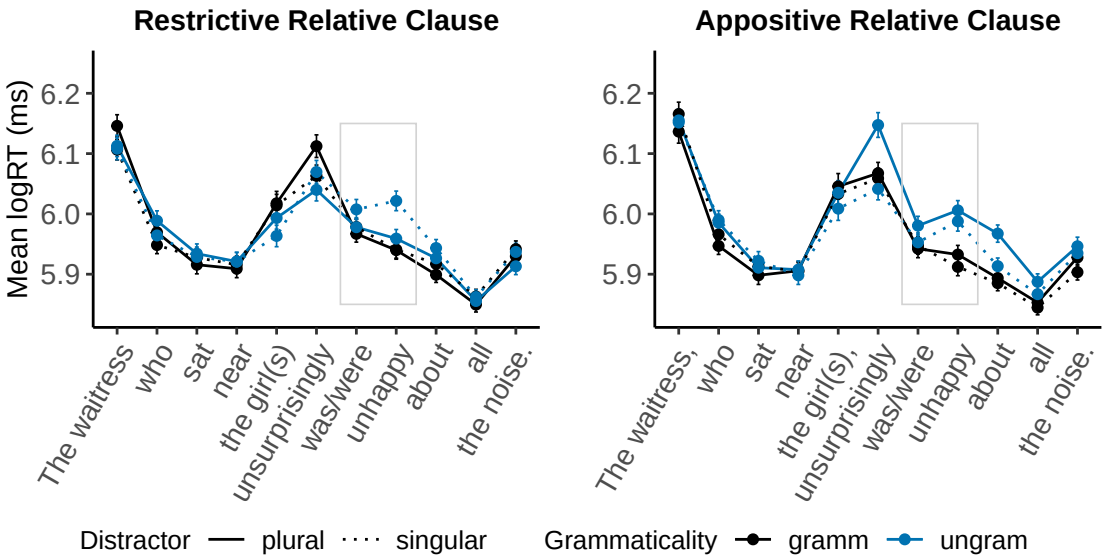


Figure 4.1: Mean log reading times in Experiment 1.
Note. Error bars represent standard errors. Regions of interest highlighted in grey boxes.

Table 4.4 presents the results of the statistical analysis with the effects of the distractor, grammaticality, and their interactions nested under each level of the predictor clause type.

Table 4.4: Summary of statistical analyses of the reading times in Experiment 1

	Regions					
	critical			spillover		
Predictors	$\hat{\beta}$	*SE*	t	$\hat{\beta}$	*SE*	*t*
(Intercept)	4.575	0.065	70.298	3.839	0.074	51.921
Clause	**0.11**	**0.005**	**2.299**	**-0.018**	**0.005**	**-3.347**
Distractor in ARC	-0.004	0.008	-0.541	-0.000	0.007	-0.064
Grammaticality in ARC	0.017	0.011	1.603	**0.067**	**0.012**	**5.523**
Distractor in RRC	0.010	0.005	1.837	**0.011**	**0.005**	**2.195**
Grammaticality in RRC	**0.018**	**0.005**	**3.316**	**0.024**	**0.005**	**4.732**
Distractor:Grammaticality in ARC	-0.009	0.011	-0.883	0.008	0.010	0.798
Distractor:Grammaticality in RRC	-0.002	0.005	-0.289	**0.010**	**0.005**	**2.013**

The analysis shows that at the critical region, there was a main effect of clause type ($\hat{\beta}$=0.11, se=0.005, t=2.299) such that RRC conditions were read longer than ARC conditions; at the spillover region, ARC conditions were read longer than RRC conditions ($\hat{\beta}$=-0.018, se=0.005, t=-3.347). Both RRCs and ARCs showed a grammaticality effect. For the RRCs, this effect appeared on both the critical ($\hat{\beta}$=0.018, se=0.005, t=3.316) and the spillover region ($\hat{\beta}$=0.024, se=0.005, t=4.732). For the ARCs, this effect appeared on the spillover region ($\hat{\beta}$=0.067, se=0.012, t=5.523), but not on the critical word ($\hat{\beta}$=0.017, se=0.011, t=1.603). Most relevant for the current purpose, at the spillover region, there was an interaction of distractor and grammaticality within the RRC condition ($\hat{\beta}$=0.010, se=0.005, t=2.013), driven by the fact that the condition with a plural distractor was read faster than the singular distractor condition but within the ungrammatical conditions. Crucially, for the ARC conditions, at the critical region, no interaction between distractor and grammaticality was found on either the critical region ($\hat{\beta}$=-0.009, se=0.011, t=-0.883) or the spillover region ($\hat{\beta}$=0.008, se=0.010, t=0.798).

Figure 4.2 illustrates the interference effect of each clause type at the critical and spillover

regions. It is only the RRC-ungrammatical condition (in the spillover region) that shows a negative interference effect, i.e., standard number agreement attraction effect.

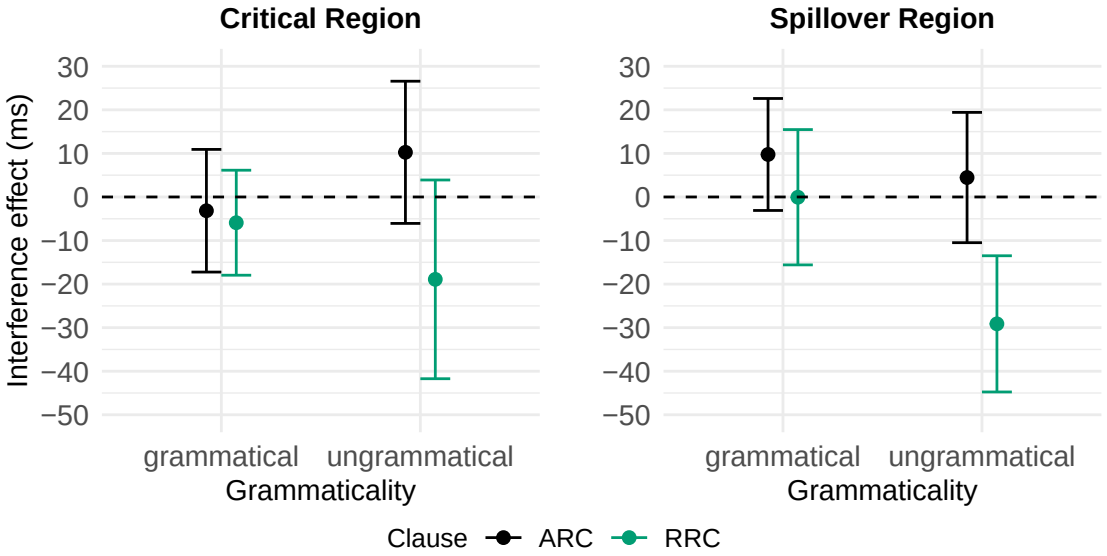


Figure 4.2: Interference effect in Experiment 1.
Note. The interference effect was calculated by subtracting the untransformed mean reading time of the singular distractor condition from the plural distractor condition. Error bars indicate 95% CI by participants.

The interference effect was calculated by subtracting the untransformed mean reading time of the singular distractor condition from the plural distractor condition, and therefore a negative value is indicative of the agreement attraction effect (as in Dillon et al. (2013) and Jäger, Engelmann, and Vasishth (2017)). There was a reliable attraction effect only at the spillover region in the ungrammatical RRC conditions.

4.3.3 Discussion

The results with the RRC condition replicated the finding in Parker and An (2018) (Experiment 1), where there was a standard number agreement attraction effect in the presence of a retrieval cue that matched the number feature of the distractor noun. The results are in line with a large body of existing findings on the cue-based retrieval mechanism in long-distance dependency resolution, evidenced by the interference effect (e.g., R. L. Lewis & Vasishth,

2005; R. L. Lewis et al., 2006; Van Dyke & McElree, 2011). Crucially, however, no reliable agreement attraction effect was found in the ARC condition. This is in line with the finding in Ng and Husband (2017) and McInnerney and Atkinson (2020), where the intervening distractor noun within a subordinate discourse unit did not interfere with the target noun during the retrieval process. The current findings are also in line with earlier work that showed a separation between ARCs and RRCs within the context of syntactic complexity (e.g., Dillon et al., 2014; Kroll & Wagers, 2019).

There are two ways to interpret the current findings. One interpretation is that the main and subordinate discourse units can be separated during memory encoding and retrieval by virtue of their distinct linguistic status. As discussed earlier, the linguistics literature has long recognized that different types of discourse units can have distinct representational statuses, with proposals suggesting that discourse units are stacked in the "attentional space" (Grosz & Sidner, 1986) one at a time in such a way that each unit can be independent of one another, or different discourse units could contribute to distinct semantic dimensions (Potts, 2005), or the subordinate discourse units are similar to speech acts that carry independent illocutionary functions distinct from the main proposition (Frazier et al., 2018; Koev, 2022). A way to implement the linguistic distinction between main and subordinate discourse status in a processing model is to assume that the parser encodes the discourse status information as features of the relevant linguistic input (e.g., [+MAIN] or [+SUBORDINATE]). These binary and static features can be used as retrieval cues to guide memory retrieval. As a result, the parser can make a distinction in real time between linguistic units based on their discourse status.

Another interpretation of the results is based on the notion of *active discourse question*. From the perspective that discourse structure is incrementally constructed, subordinate discourse information can raise active discourse questions but such questions become inactive once the subordinate discourse information is closed off. Discourse structure moves for-

ward based on which questions are raised and remain active and which questions have been resolved and can be removed. We can make a further assumption that working memory retrieval, or for the purpose of subject-verb agreement dependency at least, is sensitive to the information currently in the domain of active questions. With this approach, we can understand the results by the distractor noun inside the subordinate ARC becoming no longer accessible to memory retrieval after the closure of the ARC, at the point of retrieval. This made the target noun, which was inside the main discourse unit, become the only candidate for memory retrieval, leading to the absence of an agreement attraction effect in the ARC condition.

Both of the proposals above are consistent with the findings from Experiment 1, in which memory retrieval for an agreement-controller noun takes place after the closure of the ARC clause. However, the two proposals rely on distinct theoretical assumptions. The first proposal, which I label the *static division hypothesis*, assumes that the working memory organization of the linguistic material is sensitive to the static properties of the discourse structure such as the main vs. subordinate characteristics. The second proposal, which I label the *active question hypothesis*, assumes that the working memory organization incrementally tracks the relevant questions/issues at any given moment of an unfolding sentence. To distinguish these two proposals, in the next experiment, I modify the design such that memory retrieval of a target takes place prior to the closure of the discourse question instantiated by the subordinate ARC.

4.4 Experiment 2

In Experiment 2, the memory retrieval site is located prior to the subordinate ARC being closed off. The design of Experiments 1 and 2 are compared in (10). For the baseline control RRC conditions, we expect to observe a standard agreement attraction effect in Experiment 2, replicating the basic attraction effect in Experiment 1. As for the critical

ARC conditions, the two competing hypotheses make contrasting predictions. The *static division hypothesis* predicts that the ARC conditions in Experiment 2 would demonstrate a lack of attraction effect, similar to Experiment 1. This is because the correct retrieval target NP and the distractor NP in the ARC condition are located in different discourse units in both Experiments 1 and 2. In Experiment 1, the target NP is located in the main unit while the distractor NP is in the subordinate unit; in Experiment 2, it is the other way around.

(10) Schematization of experimental design (Experiments 1–2)

a. Experiment 1

Target(,) [who ... ***Distractor*** ...](,) **RetrievalSite** ...

b. Experiment 2

Distractor(,) [who ... **Target RetrievalSite** ...](,) ...

If the parser uses discourse status information [+MAIN] or [+SUBORDINATE] to guide retrieval, the target and distractor NP should be sufficiently distinguished from each other, reducing the probability of an interference effect. On the other hand, the *active question hypothesis* would make a different prediction. Under this hypothesis, we would expect to see an attraction effect in Experiment 2, different from Experiment 1. This is because the retrieval site in Experiment 2 is located prior to the closure of the ARC (it is inside the ARC). This means that at the point where memory retrieval of an agreement controller is initiated, the discourse questions associated with the main and the subordinate discourse units are both active, and accordingly, linguistic content in the scope of these questions is accessible for retrieval. This leads to the possibility of misretrieving the distractor, resulting in a retrieval interference effect.

4.4.1 Methods

Subjects

Experiment 2 recruited a total of 120 American English speakers residing in the U.S. (aged 18 and above) through the Prolific platform (https://www.prolific.co/). Participation recruitment was conducted with the IRB approval of the local institution. Two participants, indicating English was not their first language, were excluded, and an additional two participants who did not complete the experiment were removed, resulting in a total of 116 participants (mean age = 31.15; range: 18–50). The experiment, lasting approximately 20 minutes, provided participants with 3.50–4.00 USD in compensation.

Material and design

Material for Experiment 2 is presented in Table 4.5. The material consisted of 48 sets of items with 8 conditions in a fully crossed 2 x 2 x 2 design with grammaticality (grammatical vs. ungrammatical), distractor number (singular vs. plural), and clause type (ARC vs. RRC) as factors. I adopted the RRC material from Experiment 2 in Wagers et al. (2009). All the target sentences had the structure of *NP1(,) who NP2 VERB ..., VERB ...* (e.g., "The musician(s)(,) who the reviewer praise(e) so highly, will ..."). NP1 was the distractor noun and appeared at the matrix subject position, which was both structurally and linearly distant from the target verb. It was varied by its number feature, either a singular or plural noun (e.g., *the musician(s)*). The sentence always had an object-extracted RC structure, where NP1 was the extracted object, and NP2 was the subject in the embedded clause. NP2 was the target noun, which formed a number-agreement dependency relation with the target verb that appeared one region after the target noun and inside the RC. NP2 was always a singular noun (e.g., *the reviewer*), and the target verb either agreed with the target noun (e.g., *praises*) or not (e.g., *praise*), determining the grammaticality of the entire sentence.

ARC structures were marked with commas on the RCs. There were 12 to 20 regions for the target trial. Forty-eight grammatical filler sentences were included, which were taken from the filler items in Wagers et al. (2009).

Table 4.5: An example of the experimental item for Experiment 2

Condition	Sentence
ARC-Sg-Gr	The / *musician*, / who / the / reviewer / **praises** / **so** / highly, / will / probably / win / a / Grammy.
ARC-Sg-Ug	The / *musician*, / who / the / reviewer / **praise** / **so** / highly, / will / probably / win / a / Grammy.
ARC-Pl-Gr	The / *musicians*, / who / the / reviewer / **praises** / **so** / highly, / will / probably / win / a / Grammy.
ARC-Pl-Ug	The / *musicians*, / who / the / reviewer/ **praise** / **so** / highly, / will / probably / win / a / Grammy.
RRC-Sg-Gr	The / *musician* / who / the / reviewer / **praises** / **so** / highly / will / probably / win / a / Grammy.
RRC-Sg-Ug	The / *musician* / who / the / reviewer / **praise** / **so** / highly / will / probably / win / a / Grammy.
RRC-Pl-Gr	The / *musicians* / who / the / reviewer / **praises** / **so** / highly / will / probably / win / a / Grammy.
RRC-Pl-Ug	The / *musicians* / who / the / reviewer / **praise** / **so** / highly / will / probably / win / a / Grammy.

Note. ARC = Appositive relative clause. RRC = Restrictive relative clause. Sg = Singular distractor noun. Pl = Plural distractor noun. Gr = Grammatical. Ug = Ungrammatical. Distractor noun italicized. Regions of interest are bold-faced. The '/' sign indicates regions.

Procedure

The 48 sets of target items were pseudo-randomly assigned to eight lists under a Latin-square design. Experiment 2 also used a self-paced reading task, conducted on IbexFarm. The experiment was done in a non-cumulative word-by-word moving window fashion. Similar to Experiment 1, a 'yes/no' comprehension question task was included that targeted the content of the sentence the participants just saw (e.g., "Will the musician(s) likely win an award?"). Half of the questions targeted 'yes' and the other half 'no' as a response. Each participant read 96 sentences in total, with two-thirds of the sentences being grammatical.

No feedback on their responses was given. There were 10 practice trials before the main experiment.

Analysis

Four participants whose comprehension question accuracy was below 80% were excluded, leaving us with 112 people for the analysis. Trials with incorrect responses in the comprehension question task were removed (5.88% removal). For the self-paced reading task data analysis, the same reading time threshold was used as in Experiment 1 (3SD reading time cutoff by condition and by region), removing an additional 1.53% of the data. There were two critical regions for the main analysis: (a) the target verb region (e.g., *praise(s)*) (region 6) and (b) the spillover region (*so*) (region 7). The same analysis methods as in Experiment 1 were used.

4.4.2 Results

Comprehension accuracy

Mean comprehension question accuracy for each condition is presented in Table 4.6. The comprehension accuracy was high across all conditions.

Table 4.6: Mean comprehension question accuracy in Experiment 2

	Gr-Pl	Gr-Sg	Ug-Pl	Ug-Sg
ARC	0.944 (±0.944)	0.951 (±0.001)	0.952 (±0.001)	0.928 (±0.002)
RRC	0.958 (±0.001)	0.933 (±0.002)	0.927 (±0.002)	0.937 (±0.002)

Note. ARC = Appositive relative clause. RRC = Restrictive relative clause. Sg = Singular distractor noun. Pl = Plural distractor noun. Gr = Grammatical. Ug = Ungrammatical. Standard errors of the grand mean are in parentheses.

Reading times

Figure 4.3 shows the mean log reading times in Experiment 2.

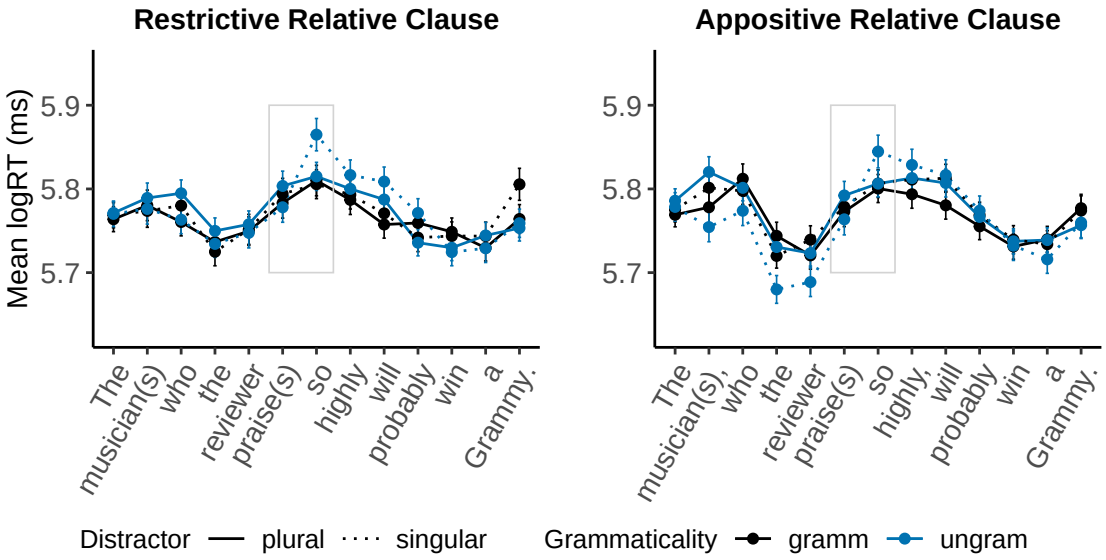


Figure 4.3: Mean log reading times in Experiment 2.
Note. Error bars represent standard errors. Regions of interest highlighted in grey boxes.

Table 4.7 reports the statistical analysis of the reading times with the effects of distractor and grammaticality nested under clause type.

Table 4.7: Summary of statistical analyses of the reading times in Experiment 2

	Regions					
	critical			spillover		
Predictors	$\hat{\beta}$	*SE*	t	$\hat{\beta}$	*SE*	*t*
(Intercept)	3.200	0.073	43.831	3.302	0.068	48.756
Clause	-0.002	0.005	-0.535	-0.003	0.005	-0.630
Distractor in ARC	-0.007	0.008	-0.899	**0.027**	**0.008**	**3.522**
Grammaticality in ARC	0.008	0.011	0.745	**0.023**	**0.011**	**2.173**
Distractor in RRC	0.000	0.005	0.067	**0.014**	**0.005**	**2.614**
Grammaticality in RRC	0.001	0.005	0.214	**0.019**	**0.005**	**3.591**
Distractor:Grammaticality in ARC	-0.011	0.011	-0.984	**0.026**	**0.011**	**2.395**
Distractor:Grammaticality in RRC	-0.002	0.005	-0.349	**0.014**	**0.005**	**2.610**

The analysis found no statistically reliable effects in the critical region. In the spillover region, however, there was a statistically reliable interaction of distractor and grammaticality both

in the ARC condition ($\hat{\beta}$=0.026, se=0.011, t=2.395) and in the RRC condition ($\hat{\beta}$=0.014, se=0.005, t=2.610).

Figure 4.4 illustrates the interference effect in each clause type at the critical and spillover regions. Both clause conditions show a negative interference effect in the ungrammatical condition in the spillover region.

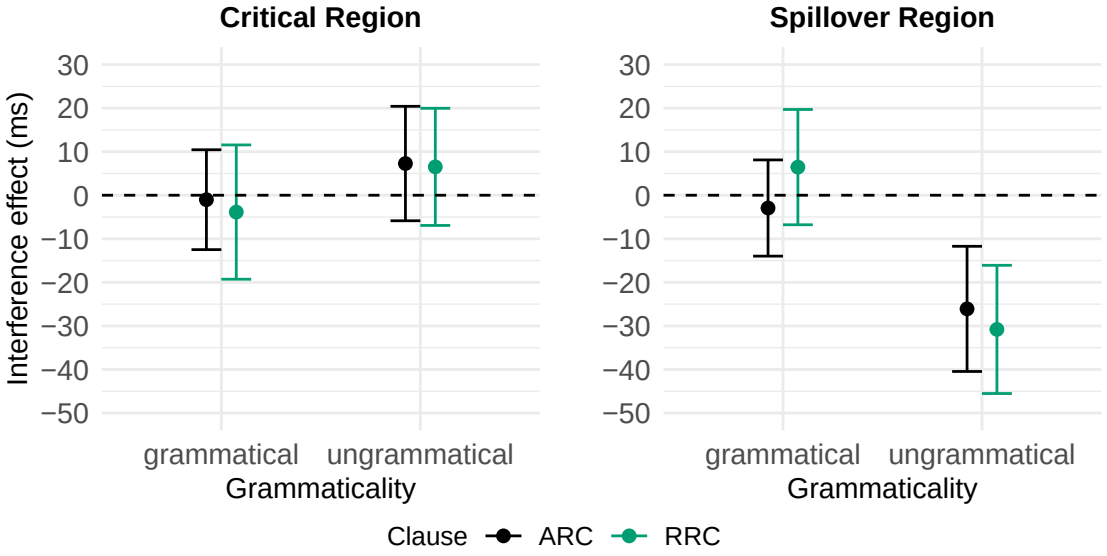


Figure 4.4: Interference effect in Experiment 2.
Note. The interference effect was calculated by subtracting the untransformed mean reading time of the singular distractor condition from the plural distractor condition. Error bars indicate 95% CI by participants.

There was an agreement attraction effect in the ungrammatical condition at the spillover region for both clause types.

4.4.3 Discussion

The agreement attraction effect in the baseline RRC conditions was observed, similar to the RRC conditions in Experiment 1. There was an attraction effect in the RRC conditions even when the distractor noun was further away from the retrieval verb than the target noun, replicating earlier findings using similar constructions (e.g., Wagers et al., 2009). More importantly, different from Experiment 1, the number agreement attraction effect was also

observed in the ARC condition. In the current experiment, the distractor inside the main discourse unit interfered with the target inside the subordinate discourse unit.

I return to the question that motivated us to conduct Experiment 2: whether the absence of an interference effect in the ARC condition in Experiment 1 was due to the linguistic units hosting the target and the distractor NPs having distinct discourse status (either [+MAIN] or [+SUBORDINATE]), or the memory retrieval being sensitive to the active state of discourse questions. The current findings lend support to the latter, supporting the active discourse question hypothesis. At the time of memory retrieval, the discourse questions associated with the linguistic units containing the target and the distractor NPs are all active, allowing the distractor to be accessible for memory retrieval and consequently resulting in interference with the target.

There is another remaining possibility, however. Let us consider the design of Experiments 1 and 2 again, as schematized in (10). The above-mentioned conclusion was based on the assumption that the two experiments differ regarding which discourse question(s) are active by the time memory retrieval of the agreement controller is initiated. However, there is another difference between the two experiments. In the design of Experiment 2 (see (10)), the distractor NP is the matrix subject NP ("The musician(s), who the reviewer praise(s)..."), and it signals an upcoming matrix verb. While the parser processes the ARC, the subject NP of the matrix clause (*the musician(s)*) is yet to be integrated with a verb. It is possible that the parser allocates additional resources to maintain the subject NP since it is part of an incomplete syntactic dependency, hence keeping the distractor NP at a relatively high activation level. The presence of the number agreement attraction effect in the ARC condition in Experiment 2, but not in Experiment 1, could be due to the fact that the distractor NP in Experiment 2 had a higher activation level than the distractor NP in Experiment 1, making it a stronger competitor to the target NP in Experiment 2. To rule out this alternative possibility, in Experiment 3 below, I changed the syntactic position of

the distractor NP to be an object NP of the matrix clause. In doing so, the distractor NP marks the end of an SVO structure instead of signaling an open dependency.

4.5 Experiment 3

I modified the material in Experiment 2 such that the distractor NP is now the object NP in the matrix clause. The core SVO argument structure at the matrix clause level is completed by the time the RC structure appears. A schematization of the manipulation is illustrated in (11):

(11) Schematization of experimental design (Experiments 2–3)

a. Experiment 2

Distractor(,) [who ... **Target RetrievalSite** ...](,) ...

b. Experiment 3

Name ... ***Distractor***(,) [who **Target RetrievalSite** ...].

In Experiment 3, similar to Experiment 2, the discourse questions associated with the target and the distractor NPs are both active when the memory retrieval of the agreement controller takes place. But different from Experiment 2, in Experiment 3, the distractor NP is an object NP and does not receive an additional activation boost resulting from the active maintenance of an incomplete dependency.

If we still observe an agreement attraction effect in the ARC conditions in Experiment 3, this would complement Experiment 2 to provide evidence for the active question-based constraint on memory retrieval.

4.5.1 Methods

Subjects

A total of 120 native speakers of American English over the age of 18 residing in the U.S. were recruited via Prolific (https://www.prolific.co/). The study was conducted with IRB approval from the local institution. The duration of the experiment was about 20 minutes, and participants were paid 3.50–4.00 USD in compensation. Four participants whose self-reported first language was not English were removed. One participant did not complete the experiment. This left us with 115 participants (mean age = 33.37; range: 18–50).

Material and design

Table 4.8 shows an example of material used for Experiment 3. There were 48 sets of 8 items, with 8 conditions, varied by grammaticality (grammatical vs. ungrammatical), distractor number (singular vs. plural), and clause type (ARC vs. RRC). The same material as in Experiment 2 was used, but it was modified in a way that the structure had an SVO structure, followed by an ARC or an RRC. The target material had the structure of *Name VERB NP1(,) who NP2 VERB ...* (e.g., "Alicia met the musician(s)(,) who the reviewer praise(s) so highly."), where ARCs were marked with a comma. Forty-eight different names (in the *Name* position) were used for the matrix subject. The number of regions for the target trial ranged from 12 to 14. The same filler sentences in Experiment 2 were used.

Procedure

The 48 target trials, along with 48 filler sentences, which were all grammatical, were distributed across eight lists with a Latin-square design. Each participant was assigned 96 sentences in total. The ratio of grammatical to ungrammatical sentences in the trials was 2:1. There were 13 practice trials before the main trials. The procedure was the same as

Table 4.8: An example of the experimental item for Experiment 3

Condition	Sentence
ARC-Sg-Gr	Alicia / met / the / *musician*, / who / the / reviewer / **praises** / **so** / highly.
ARC-Sg-Ug	Alicia / met / the / *musician*, / who / the / reviewer/ **praise** / **so** / highly.
ARC-Pl-Gr	Alicia / met / the / *musicians*, / who / the / reviewer / **praises** /**so** / highly.
ARC-Pl-Ug	Alicia / met / the / *musicians*, / who / the / reviewer / **praise** / **so** / highly.
RRC-Sg-Gr	Alicia / met / the / *musician* / who / the / reviewer / **praises** / **so** / highly.
RRC-Sg-Ug	Alicia / met / the / *musician* / who / the / reviewer / **praise** / **so** / highly.
RRC-Pl-Gr	Alicia / met / the / *musicians* / who / the / reviewer / **praises** /**so** / highly.
RRC-Pl-Ug	Alicia / met / the / *musicians* / who / the / reviewer / **praise** / **so** / highly.

Note. ARC = Appositive relative clause. RRC = Restrictive relative clause. Sg = Singular distractor noun. Pl = Plural distractor noun. Gr = Grammatical. Ug = Ungrammatical. Distractor noun italicized. Regions of interest are bold-faced. The '/' sign indicates regions.

in Experiment 2. Each trial was followed by a comprehension question. Filler trials had a 'yes/no' comprehension question task that asked a question related to the content of the sentence. Target trials had a forced-choice comprehension question task that probed participants' understanding of the *wh*-pronoun (e.g., "Who does the reviewer praise highly?"). Participants were then asked to select one of the two options, (a) *Alicia* or (b) *the musician(s)*. The order of the choices was randomized.

Analysis

Eleven participants were excluded from the main analysis based on the comprehension question accuracy (below 80%), leaving us with data points from 104 participants for the main analysis. For the reading time result analysis, only the responses that selected the object (e.g., *the musician(s)*) (as opposed to the matrix subject, e.g., *Alicia*) in the question were included for the main data analysis (3.09% data removal). Additionally reading times beyond 3SD by condition and by region were removed (additional 1.46% data removal). Two regions were identified as regions of interest for the analysis: (a) the target verb region (*praise(s)*) (region 8) and (b) the spillover region (*so*) (region 9). The same analysis methods as in previous experiments were used.

4.5.2 Results

Referent selection accuracy

Table 4.9 illustrates the mean proportion of selecting the object (rather than the matrix subject) as the antecedent of the *wh*-pronoun. The selection task results indicate a strong tendency among participants to parse the object of the matrix clause (e.g., *the musician(s)*) as the object of the embedded verb (e.g., *praise(s)*).

Table 4.9: Mean proportion of selection of the object in Experiment 3

	Gr-Pl	Gr-Sg	Ug-Pl	Ug-Sg
ARC	0.987 (±0.001)	0.976 (±0.002)	0.962 (±0.002)	0.966 (±0.002)
RRC	0.969 (±0.002)	0.960 (±0.002)	0.968 (±0.002)	0.963 (±0.002)

Note. ARC = Appositive relative clause. RRC = Restrictive relative clause. Sg = Singular distractor noun. Pl = Plural distractor noun. Gr = Grammatical. Ug = Ungrammatical. Standard errors of the grand mean are in parentheses.

Reading times

Mean log reading times for Experiment 3 are presented in Figure 4.5.

Table 4.10 shows the statistical analysis of reading time results.

Table 4.10: Summary of statistical analyses of the reading times in Experiment 2

	Regions					
	critical			spillover		
Predictors	$\hat{\beta}$	*SE*	t	$\hat{\beta}$	*SE*	*t*
(Intercept)	3.841	0.073	52.979	3.310	0.075	44.045
Clause	0.002	0.005	0.406	**-0.011**	**0.005**	**-2.463**
Distractor in ARC	-0.009	0.008	-1.196	**0.024**	**0.007**	**3.279**
Grammaticality in ARC	0.016	0.011	1.454	**0.024**	**0.010**	**2.322**
Distractor in RRC	0.004	0.005	0.808	0.010	0.005	1.952
Grammaticality in RRC	-0.005	0.005	-0.932	**0.015**	**0.007**	**2.328**
Distractor:Grammaticality in ARC	0.007	0.011	0.695	**0.026**	**0.010**	**2.506**
Distractor:Grammaticality in RRC	-0.001	0.005	-0.203	0.008	0.005	1.558

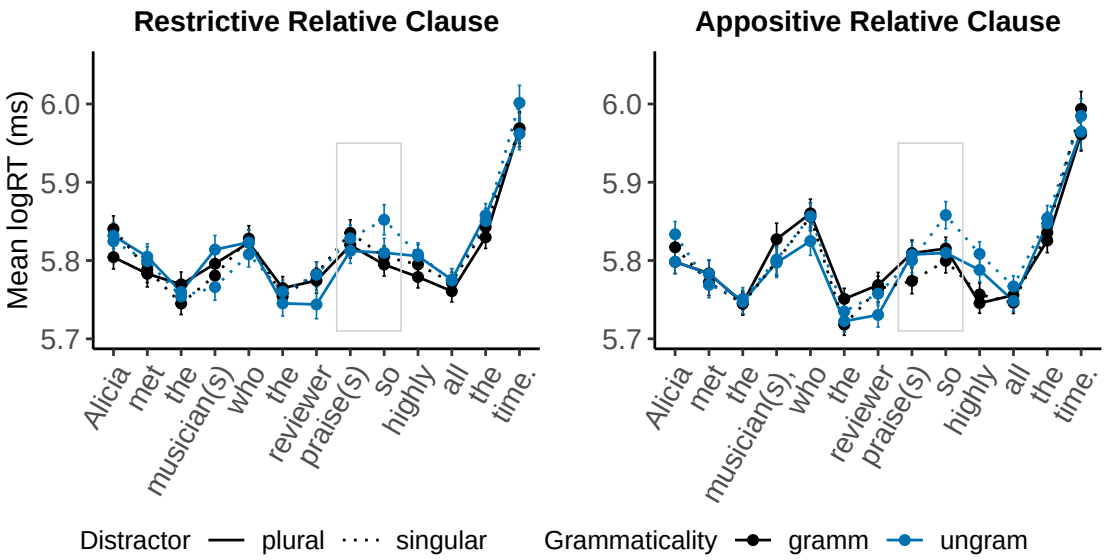


Figure 4.5: Mean log reading times in Experiment 3.
Note. Error bars represent standard errors. Regions of interest highlighted in grey boxes.

No statistically reliable effects were found in the critical region. In the spillover region, there was an interaction of distractor and grammaticality in the ARC condition ($\hat{\beta}$=0.026, se=0.010, t=2.506), indicating an agreement attraction effect. There was a main effect of grammaticality ($\hat{\beta}$=0.015, se=0.007, t=2.328) and a marginal effect of distractor ($\hat{\beta}$=0.010, se=0.005, t=1.952) in the RRC condition. However, no interaction of distractor and grammaticality was found in the RRC condition ($\hat{\beta}$=0.008, se=0.005, t=1.558).

The mean interference effects are presented in Figure 4.6. Both ARC and RRC conditions show a negative interference effect in the ungrammatical condition in the spillover region.

4.5.3 Discussion

The number agreement attraction effect was observed in the ARC conditions in Experiment 3, where the distractor NP was not part of an incomplete dependency. The results of Experiments 2 and 3 together provide evidence for the hypothesis that information within main and subordinate discourse units can interfere with each other during memory retrieval when they are both associated with active discourse questions. It is not entirely clear to us

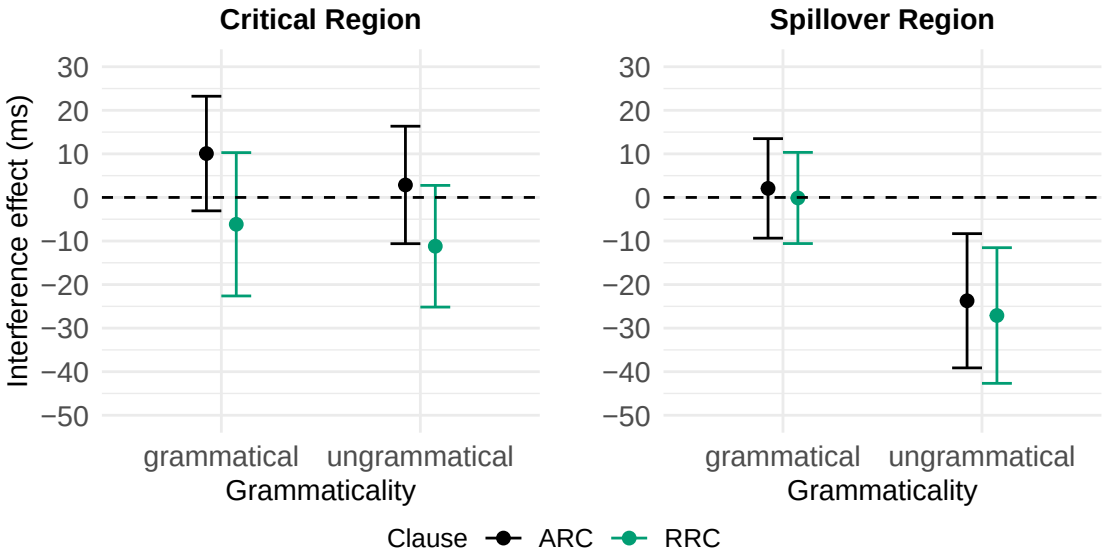


Figure 4.6: Interference effect in Experiment 3.

Note. The interference effect was calculated by subtracting the untransformed mean reading time of the singular distractor condition from the plural distractor condition. Error bars indicate 95% CI by participants.

why there was not a reliable agreement attraction effect with the RRC conditions in Experiment 3. There was a numerical trend of a standard agreement attraction effect in the RRCs, with an effect size of around 30 ms (Figure 4.6), which is in line with the effect sizes reported in previous studies (Jäger et al., 2017, for a meta-analysis review). It is possible that there was inadequate power in the current experiment. It is also worth noting that some previous work has shown similar results indicating that retrieval interference effects are weaker when the distractor is an object rather than a subject (Parker & An, 2018; Van Dyke & McElree, 2011), possibly due to reduced prominence of object NPs relative to subject ones.

4.6 General discussion

4.6.1 Summary of the findings

Across three experiments, the canonical standard number agreement attraction effect was found in the RRC conditions. Meanwhile, the effect was absent in the ARC condition

in Experiment 1. However, the attraction effect was present even in the ARC condition in Experiments 2 and 3. This is summarized in Table 4.11, along with the two initial hypotheses.

Table 4.11: Summary of the findings in Experiments 1–3 factored by discourse status and active state of discourse questions

Exp.	Clause	Discourse status [Target, Distractor]	Active state of discourse questions [Target, Distractor]	Number agreement attraction
1	RRC	[+ MAIN, + MAIN]	[+ ACTIVE, + ACTIVE]	yes
	ARC	[+ **MAIN**, + **SUBORD.**]	[+ **ACTIVE**, −**ACTIVE**]	**no**
2 & 3	RRC	[+ MAIN, + MAIN]	[+ ACTIVE, + ACTIVE]	yes
	ARC	[+ **SUBORD.**, + **MAIN**]	[+ **ACTIVE**, + **ACTIVE**]	**yes**

Note. Subord. = Subordinate

The results collectively suggest that the observed findings are best predicted under the active state of discourse questions of the retrieval target and the distractor. The differing discourse status of the target or the distractor (i.e., one being part of the main and the other the subordinate discourse structure) does not make the correct predictions.

4.6.2 Between-experiment comparison

Since the contrast between experiments is crucial to compare the hypotheses, I conduct a between-experiment comparison specifically targeting the presence and absence of the number agreement attraction effect. I present two additional analyses for this direct comparison.

First, I analyzed the log-transformed reading time data only from the ARC conditions at the spillover region. A mixed-effects linear regression model was built with fixed effects of grammaticality (grammatical vs. ungrammatical), distractor number (singular vs. plural), experiment (Experiments 1–3), and their interactions; word length was also included as a fixed effect to account for the lexical variation across experiments. By-participant and by-item random intercepts were also included. The grammaticality and distractor variables were sum-coded (grammatical = −0.5; ungrammatical = 0.5; plural = −0.5; singular = 0.5).

For the experiment variable, since the key observation in the current work comes from the contrast between Experiment 1 vs. Experiments 2 and 3, a Helmert effect coding scheme was used to compare Experiment 1 with Experiments 2 and 3. The Helmert coding also allowed us to compare Experiment 2 with Experiment 3. The summary of the model is shown in Table 4.12.

Table 4.12: Summary of statistical analyses of the reading times across experiments in the ARC condition at the spillover region

	Estimate	Std. Error	t value
(Intercept)	5.818	0.019	314.200
Distractor	0.010	0.007	1.447
Grammaticality	**0.041**	**0.007**	**6.246**
Experiment(1 vs. 2–3)	**0.113**	**0.035**	**3.237**
Experiment(2 vs. 3)	**-0.014**	**0.060**	**-0.237**
Distractor:Grammaticality	**0.035**	**0.013**	**2.650**
Distractor:Experiment(1 vs. 2–3)	**-0.029**	**0.014**	**-2.056**
Distractor:Experiment(2 vs. 3)	0.014	0.024	0.591
Grammaticality:Experiment(1 vs. 2–3)	**0.049**	**0.014**	**3.485**
Grammaticality:Experiment(2 vs. 3)	-0.004	0.024	-0.154
Distractor:Grammaticality:Experiment(1 vs. 2–3)	-0.043	0.028	-1.517
Distractor:Grammaticality:Experiment(2 vs. 3)	-0.031	0.048	-0.637

Although the variables distractor and grammaticality each had an interaction with the contrast Experiment 1 vs. 2–3 (Distractor:Experiment $\hat{\beta}$=-0.029, se=0.014, t=-2.056; Grammaticality:Experiment $\hat{\beta}$=0.049, se=0.014, t=3.485), there was no reliable three-way interaction ($\hat{\beta}$=-0.043, se=0.028, t=-1.517). It is possible that there was not sufficient power for a three-way interaction to emerge.

As the agreement attraction effect was observed only in ungrammatical sentences, I also conducted a second analysis focusing exclusively on the ungrammatical ARC conditions. A mixed-effects model was built similar to the one in the previous analysis, but removing the grammaticality variable (Table 4.13).

The analysis showed a statistically reliable interaction between distractor and experiment, specifically the distractor effect being modulated by the difference between Experiment 1 vs.

Table 4.13: Summary of statistical analyses of the reading times across experiments in the ARC ungrammatical condition at the spillover region

	Estimate	Std. Error	t value
(Intercept)	5.835	0.021	280.543
Distractor	**0.026**	**0.010**	**2.650**
Experiment(1 vs. 2–3)	**0.136**	**0.037**	**3.657**
Experiment(2 vs. 3)	-0.016	0.063	-0.251
Distractor:Experiment(1 vs. 2–3)	**-0.050**	**0.021**	**-2.367**
Distractor:Experiment(2 vs. 3)	-0.002	0.036	-0.064

Experiments 2 and 3 ($\hat{\beta}$=-0.050, se=0.021, t=-2.367), driven by the fact that the number agreement attraction effect in the ARC condition in Experiment 1 differed from the effect in Experiments 2–3.

The second analysis in particular strengthens the current finding that the retrieval interference effect in the ARC condition patterns similarly in Experiments 2 and 3 but differs from Experiment 1. In Section 4.6.3, a more detailed account is provided of the way the active state of discourse questions plays a role in the retrieval process.

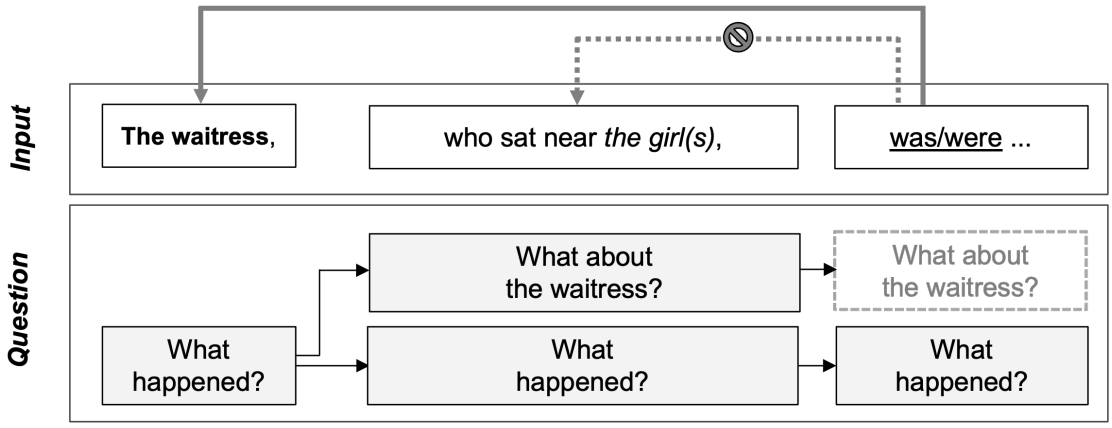
4.6.3 Discourse questions and retrieval

Given the alignment with the active state of discourse questions as our primary hypothesis, I use a question-based approach (Jasinskaja, 2016; Riester, 2019) to model how discourse structure is incrementally constructed and how that incremental process exerts an effect on the memory access of representations with different discourse statuses. In particular, it is assumed that a discourse structure can be represented as a question stack. Different discourse issues/questions arise while a discourse progresses, and these questions stay active until they are resolved and removed from the question stack. In our proposal, representations associated with active discourse questions are more accessible for memory retrieval, while representations associated with inactive questions are less accessible. This could be the case because active questions represent issues that the current discourse is engaged with, and

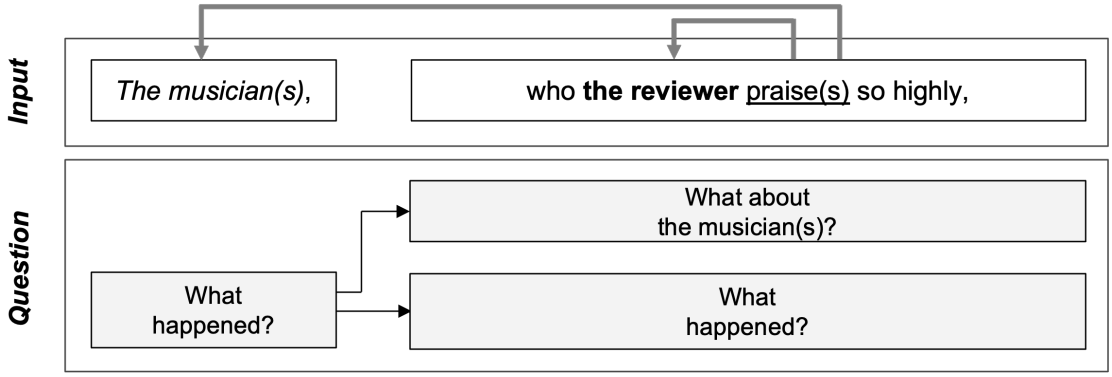
therefore more attention could be allocated to information associated with active questions. On the other hand, information that is not immediately relevant for resolving the active questions is less activated and less accessible. This is on par with the proposal in Grosz and Sidner (1986), where the authors suggested that "information about the objects, properties, relations, and discourse intentions that are most salient" in discourse structure becomes part of the attentional state, but anything that is irrelevant to the discourse development will be popped off the focus space. This idea is also broadly in line with previous work showing that active questions or Questions under Discussion (QUDs) can introduce focus to specific aspects of a target sentence and facilitate deeper processing (Clifton & Frazier, 2018; Cutler & Fodor, 1979; Sanford, 2002; Sanford & Sturt, 2002). For instance, Clifton and Frazier (2018) showed that sentences related to the QUD are processed faster. Similarly, Cutler and Fodor (1979) found that introducing a question before the target material helps comprehenders to focus on the relevant information.

In Section 3.2.1, I schematically illustrated how "active question" can be used to understand the discourse status of RRCs and ARCs (see Figures 3.1 and 3.2. In Figure 4.7, I illustrate how the difference between the ARC conditions in Experiments 1 and 2 in their active state of discourse questions accounts for the retrieval process.

Figure 4.7a illustrates the ARC conditions in Experiment 1. Two discourse questions are created and processed incrementally. One is associated with the matrix clause ("What happened?") and the other is associated with the ARC ("What about the waitress?"). The ARC-associated question is removed from the stack after the ARC ends. Memory retrieval for the agreement controller starts at the matrix verb (*was/were*), and by then, the ARC-related discourse question is inactive. This leads to reduced accessibility to representations associated with the ARC, including the distractor NP (*the girl(s)*), hence reducing the likelihood of an agreement attraction effect. This is in contrast with Experiments 2, as illustrated in Figure 4.7b. Again, two discourse questions are incrementally constructed, with one as-



(a) ARC in Experiment 1



(b) ARC in Experiment 2

Figure 4.7: Discourse questions and retrieval in the ARC condition in Experiments 1–2. *Note.* The box with dotted lines indicates the discourse question is no longer active. The linguistic inputs pointed with solid arrows indicate information accessible for retrieval; those pointed with dotted arrows indicate information inaccessible for retrieval. Bold-faced word = target word; italicized word = distractor; underlined word = retrieval site.

sociated with the matrix clause and the other associated with the ARC. Since the memory retrieval of the agreement controller is initiated at a moment before the ARC is completed (i.e., at the verb *praise(s)*), the target NP (*the reviewer*) is associated with an active question ("What about the musicians?"). Also importantly, at the moment of memory retrieval, the distractor NP (*the musician(s)*) is also situated within an active discourse question (i.e., the question associated with the matrix clause "What happened?"). Both the target agreement controller (*the reviewer*) and the distractor (*the musician(s)*) are accessible for memory retrieval, resulting in a standard agreement attraction effect. The ARC condition in Exper-

iment 1 showed no agreement attraction effect, consistent with the fact that it is exactly in these conditions that the distractor NPs are associated with a [−ACTIVE] discourse question. The difference between the ARC conditions in Experiments 1 and 3 can be captured in very similar ways, as Experiments 2 and 3 have the same configuration of the active state of discourse questions between the target and the distractor.

Discourse questions and the event model

Under our proposal, the [+ACTIVE] and [−ACTIVE] discourse questions essentially partition the discourse into distinct "domains," which results in different levels of accessibility for memory retrieval. This idea overlaps in some interesting ways with the idea of an "event" entity in Event Segmentation Theory (e.g., Radvansky & Zacks, 2017; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). Under this framework, our cognitive system segments activities into events and represents the world around us using various event models. The event models are connected through causal or other types of logical relations. The specific ways these events and event models are constructed have consequences for how we perceive and remember our experience. For example, information from the current event model is more accessible than information from previous events that are not the focus of the current attention. Therefore, it could be more difficult to retrieve information from an event after having passed the relevant event boundary.

Event segmentation has been shown to affect narrative comprehension (Bailey, Kurby, Sargent, & Zacks, 2017; Smith, Kurby, & Bailey, 2023; Zacks, Speer, & Reynolds, 2009). But it is an open question of what specific linguistic features guide people's segmentation of linguistic events. I suggest that discourse questions are potentially useful constructs to help delineate event boundaries in language comprehension. An [+ACTIVE] discourse question is an issue that the current discourse is engaged with, establishing salient boundaries from those [−ACTIVE] discourse questions. Information within the currently activated event model,

encompassed by the [+ACTIVE] discourse question, is highly available for memory retrieval, and potentially interfere with each other. But an [−ACTIVE] discourse question, similar to an inactive (or less activated) event, would be less accessible for memory retrieval.

4.6.4 Alternative approaches

In addition to our question-based account, several alternative accounts could potentially explain the findings of the current study. These include factors such as rapid loss of representation of the subordinate discourse structure, distinct prosodic contour, and inherently different parsing mechanisms. These will be discussed in the following four sections. I illustrate how these alternative accounts may not be sufficient to explain the current findings.

Rapid loss of syntactic structure representations

Dillon et al. (2017) suggested that the syntactic form of the appositive structures, in contrast to regular relative clauses, can be lost rapidly, making it less accessible for memory retrieval. The idea of syntactic loss of appositive structure has some appeal if we assume that the appositive content will eventually become part of the given information in a comprehender's mental model (e.g., AnderBois et al., 2015), with the conceptual representation of the appositive content held in memory, while the precise structural form decays quickly.

The drawback of this account, however, is that the empirical evidence is mixed regarding whether surface structures indeed undergo a rapid forgetting process. Some work has shown evidence that syntactic forms decay fairly quickly (Lombardi & Potter, 1992; Potter & Lombardi, 1990). For instance, Lombardi and Potter (1992) examined the robustness of maintaining the surface form of a sentence. In a sentence-recall paradigm, participants were presented with a target sentence and were instructed to recall the sentence. In between the exposure and recall phrases, participants were also presented with a distractor verb that was synonymous with the verb in the target sentence. It was found that even the brief exposure

to a simple distractor verb was sufficient to disrupt recall accuracy, and this provided some evidence that the surface form of a sentence is short-lived. On the other hand, there is also work showing that surface forms may last for a long time (Gurevich et al., 2010; Kaschak et al., 2011). For instance, Kaschak et al. (2011) found a strong structural priming effect with a long lag (even as long as a week) between the priming exposure and the target production task. Participants in their study showed a strong tendency to repeat the structure they were exposed to even after a week from the exposure phase. It is difficult to directly compare these studies since they have used different paradigms. But it is fair to conclude that the specific constraints on when and how structural forgetting could take place are yet to be established.

The effect of prosody

Another promising approach for understanding the unique properties of the ARCs highlights the role of prosody. ARCs in English, as opposed to RRCs, are commonly marked with an intonation boundary before the RC, usually signaled with a pause (Dehé, 2014; Watson & Gibson, 2004). Earlier work has shown that prosodic boundaries in general can impact parsing decisions, such as resolving ambiguities, and guiding syntactic attachment preferences (Breen, 2014; Fodor, 1998, 2002; Frazier, Carlson, & Clifton, 2006; Jun & Bishop, 2015; Wagner & Watson, 2010; Watson & Gibson, 2004) and mitigating or intensifying processing difficulty (Bader, 1998). Although our testing stimuli were presented visually, we cannot exclude the effects of implicit prosody (Fodor, 1998, 2002). Previous work has shown that grouping sentences into prosodic chunks facilitates the processing of the upcoming linguistic material (Hirotani, Frazier, & Rayner, 2006; Staub, 2007), indicative of a close relationship between prosodic grouping and syntactic parsing. Particularly relevant for the current purpose, Hirotani et al. (2006) found that ARCs are read faster than the RRC counterparts, which they interpreted as a facilitation effect of integrating the upcoming linguistic mate-

rial with the help of intonational grouping. Along similar lines, Kroll and Wagers (2019) suggested that prosodic grouping and syntactic parsing can go in tandem.[4] Syntactic constituents are less likely to compete for short-term memory resources when they are in separate prosodic groupings. Appositives, therefore, are less likely to compete with the hosting matrix clause, given a clear prosodic boundary from the matrix clause, in contrast to RRCs.

The connection between prosodic grouping and syntactic parsing appears valid, and it is likely at work while processing ARCs. However, this approach alone is not sufficient to account for all the results in the current study. In particular, I note that in all of our experiments, there was a prosodic difference between ARCs and RRCs. The prosodic grouping account will not be able to explain why the prosodic differences between ARCs and RRCs only led to different agreement attraction effects in Experiment 1, but not in Experiments 2–3. The discourse question-based account was developed to capture the differences between the three experiments. It is entirely possible that prosodic grouping and discourse-question tracking are two independent processes that are simultaneously at work. Whether there could be interactions between these two processes is an avenue for future work.

Dependency length difference between experiments

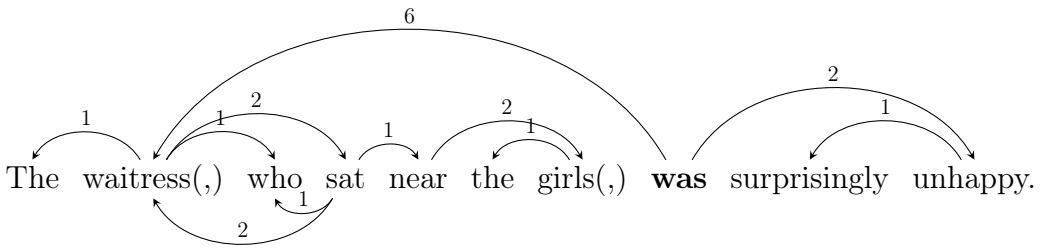
The structures I tested in Experiments 1–3 are not entirely the same. This leads to the potential concern that different parsing mechanisms may have been evoked in these experiments, which can independently explain the observed effects. One potential structural difference across different experiments is dependency length. Experimental work and large-scale corpus studies have shown that languages exhibit preferences for minimizing dependency length (Futrell, Mahowald, & Gibson, 2015; Gibson et al., 2019; Gildea & Temperley, 2007, 2010;

4. The goal of Kroll and Wagers (2019) was to identify the underlying difference that leads to the reduced processing cost in sentences with embedded appositives compared to RRCs. One of the tested hypotheses was the role of QUD, and the authors manipulated the QUD (Experiments 2–3). In one condition the question targeted only the main content whereas the other contained a coordinated question, where one targeted the main and the other the not-at-issue appositive content. No reliable effects of QUD were found.

Liu, 2008; Liu, Xu, & Liang, 2017; Temperley & Gildea, 2018; Yadav, Mittal, & Husain, 2022). Dependency length minimization (DLM) has been also correlated with cognitive constraints such as limited working memory capacity in a way that sentences with a longer dependency incur greater processing cost than those with shorter dependency (e.g., Gibson, 1998; Grodner & Gibson, 2005). It is possible that structures with a larger count of dependency length may be more susceptible to interference effect, due to higher demand of working memory resources.

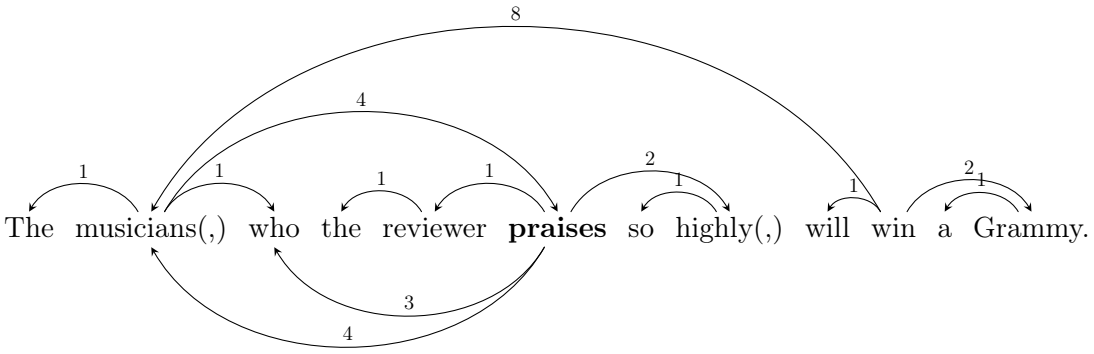
I calculated the total dependency length of the structures tested in Experiments 1 and 2. Following the Universal Dependencies (UD) project guidelines (version 2) (Nivre et al., 2020) and adopting the method from Futrell et al. (2015), I measured the total dependency length from the beginning of the target sentence up till the critical retrieval site, where the number agreement dependency is completed (see Figure 4.8). The dependency length was calculated by summing all the closed dependencies on the left-hand side of the retrieval site. Most importantly, the RRC and ARC structures I tested do not differ in their total count of dependency length.

The summed dependency length until the retrieval site is 17 in Experiment 1 (Figure 4.8a), and 15 in Experiment 2 (Figure 4.8b). The difference between the two experiments is small. If a slightly higher count of total dependency length could nonetheless be more demanding on working memory and potentially result in a higher degree of memory interference, we might have expected more agreement attraction effect in Experiment 1 than in Experiment 2. This prediction was not borne out. The two experiments exhibited a similar number agreement attraction effect in the RRC conditions; and for the ARCs, it was Experiment 2 instead of Experiment 1 that exhibited an attraction effect.



(a) Dependency in Experiment 1

Note. Total dependency length until the retrieval site (*was*) = 17.



(b) Dependency in Experiment 2

Note. Total dependency length until the retrieval site (*praises*) = 15.

Figure 4.8: Dependency length calculation.

Different position of the distractor

Another structural difference between Experiment 1 and Experiments 2–3 concerns the position of the distractor. In Experiment 1, the distractor intervenes between the subject-verb agreement dependency, but in Experiments 2–3, the distractor does not intervene. Staub (2010) argued that depending on the position of the distractor, there may be different types of number agreement attraction effects. In particular, in Staub (2010), the agreement attraction effect in a distractor-non-intervening construction was mostly driven by trials with long RTs, i.e., the effect was primarily present in the right tail of the RT distribution. On the other hand, in a distractor-intervening construction, the agreement attraction effect was more evenly distributed in the entire RT distribution. This difference was considered to

be indicative of a qualitatively distinct processing mechanism underlying the agreement attraction effect in different constructions. If intervening and non-intervening distractors are associated with distinct parsing mechanisms, this could potentially challenge our discourse question account of the current findings.

I created vincentile plots to take a closer look into the data points that led to the interference effect (Figure 4.9). The vincentile plots were generated following the procedures outlined in Staub (2010). Focusing on the conditions most relevant for the agreement attraction effect, Figure 4.9 only targeted the ungrammatical conditions at the spillover region. For each clause type (RRC or ARC), I divided the data points for each subject into vincentiles, with the fastest 10% of RTs as vincentile 1, the next fastest group as vincentile 2, and so on. Subsequently, the agreement attraction effect was computed by subtracting the singular ungrammatical condition from the plural ungrammatical condition for each subject and each vincentile. Then an averaged agreement attraction effect was calculated for each vincentile. The Y-axis values smaller than 0 ms in Figure 4.9 indicate a standard number agreement attraction effect.

As shown by the results for the RRC conditions (left panel in Figure 4.9), the size of the agreement attraction effect was much larger in the slower RT range, suggesting that the skewness difference instead of the mean difference between the plural vs the singular distractor condition is more responsible for the observed attraction effect. Most crucially, the pattern appears consistent across all three experiments, suggesting that the different positions of the distractor did not necessarily lead to distinct processing mechanisms in the current study. For the ARC conditions (right panel in Figure 4.9), consistent with the results reported in the result section, Experiment 1 did not show a robust agreement attraction effect. Furthermore, Experiments 2–3, both containing non-intervening distractors, showed similar patterns. Much like the agreement attraction effects in the RRCs, the attraction effect in the ARCs was most prominent in the slower RT range. These findings suggest that

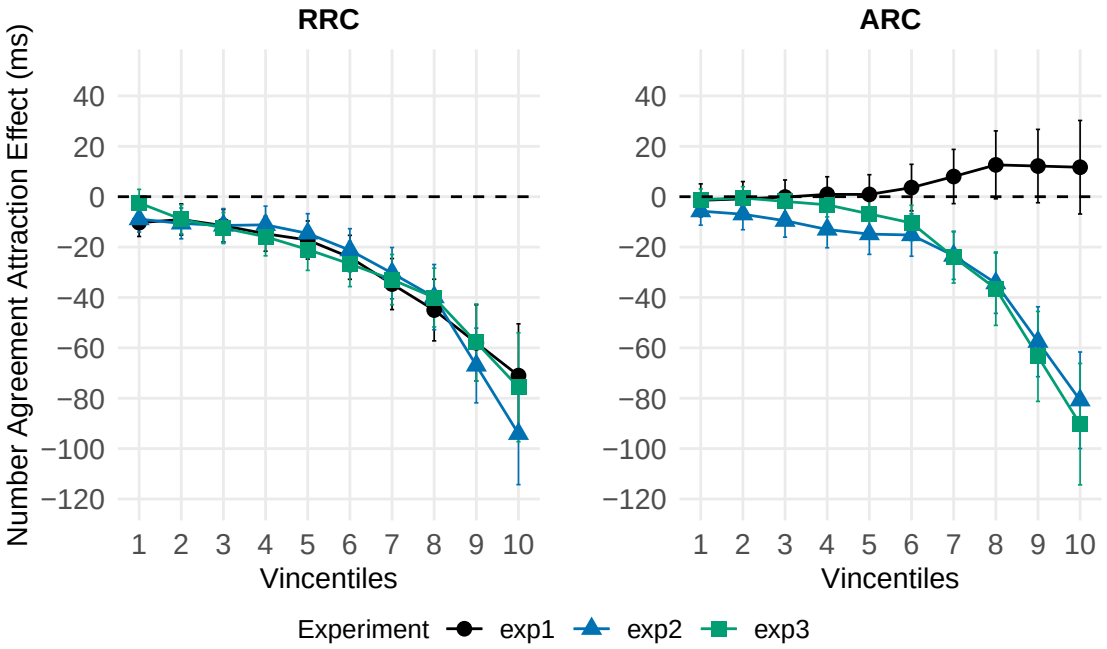


Figure 4.9: Vincentile plot of number agreement attraction effect in Experiments 1–3.
Note. The effect was calculated by subtracting the ungrammatical singular condition from the ungrammatical plural condition. Error bars represent standard errors.

the variations in configuration between the target and distractor across experiments do not fully account for the distinct interference effect observed between experiments.

I also used a statistical approach to understand the distribution of the data. An ex-Gaussian analysis was conducted on the observed agreement attraction effects, following the methods in Staub (2010) and K.-J. Huang and Dillon (2023). Ex-Gaussian modeling allows us to estimate whether the agreement attraction effects observed in Experiments 1–3 reside in the shift of the mean between the RT distribution of the target condition and its control condition, or by a shift in skewness (i.e., an effect in the tail of the RT distribution), or a combination of both. More importantly, since the RRC conditions (in particular from Experiments 1 and 2) in the current study were similar to the stimuli used in Staub (2010), I will use the results from the RRC conditions to evaluate whether the agreement attraction effects in the current study show qualitatively distinct patterns based on the intervening and non-intervening distractor contrast discussed in Staub (2010). Furthermore, for a comparison

between RRC and ARC conditions, I also fitted ex-Gaussian models to the ARC data.

To examine whether the agreement attraction effects observed in Experiments 1–3 were driven by the mean difference between the target condition and its control condition, or by a shift in skewness (i.e., an effect from the tail of the distribution), or a combination of both, I fitted ex-Gaussian models to the data. Ex-Gaussian modeling allows us to separately estimate effects from the shift of the mean RT (the μ parameter) and effects from the shift of skewness (the τ parameter). Linear mixed-effects models with an ex-Gaussian distribution were implemented in R (R Core Team, 2022) using the `brms` library (Bürkner, 2017). As our primary interest was the number agreement attraction effect, which only appeared in the ungrammatical conditions, we analyzed the data from the ungrammatical conditions at the spillover region only. A separate model was fit for each clause type and experiment, where the raw reading time (ms) was the dependent variable, and the distractor was included as the fixed effect (treatment coded singular = 0; plural = 1). Random effects of item and subject were also included in the model. The ex-Gaussian distribution had 3 parameters, μ (mean), τ (skewness), and σ (standard deviation). It was assumed that there was no distractor effect on σ.

Tables 4.14–4.16 present the model results, the subset data points associated with the number agreement attraction effect fitted to the ex-Gaussian distribution.

Ex-Gaussian distribution fit to data in Experiment 1 I start with the RRC condition in Experiment 1 (Table 4.14a). The intercept is the grand mean of the singular condition (= 430.46 ms). The "Distractor" is the mean reading time difference between the two distractor conditions (=−25.71 ms); the mean reading time of the plural condition is 416.12 ms (= 441.83+ (−25.71)). The skewness of the data points for the singular distractor condition (τ) is around 132.95 ms (= exp(4.89)), and around 105.64 ms (= exp(4.89+ (−0.23))) for the plural distractor condition. The distractor effect on τ can then be calculated by subtracting the τ effect on the singular condition from the τ effect on the plural condition, which amounts

to -27.31 ms (= 105.64 ms−132.95 ms). Finally, the distractor effect on μ is calculated by subtracting the τ effect from the mean difference between the two distractor conditions, which amounts to 1.6 ms (=−25.71−(−27.31)). In summary, the key observation is that the mean difference between the two distractors is **−25.71 ms**, and this effect, i.e., the number agreement attraction effect, is driven by data points mainly from the effect of τ (= **−27.31 ms**) and trivially from the effect of μ (= **1.6 ms**). I compare this distribution of the data points with that in Experiment 1 ARC condition (Table 4.14b). The grand mean of the reading time concerning the baseline, singular condition is 430.46 ms. The effect of Distractor is **−2.17 ms**; hence, the mean reading time of the plural condition is 428.29 ms (= 430.46+ (−2.17)). The skewness in the singular condition is approximately 115.58 ms (= exp(4.75)), and 112.17 ms (= exp(4.75+ (−0.03)) in the plural condition. Based on these values, I calculate the distractor effect on τ: approximately **3.41 ms** (= 115.58−112.17). Finally, the distractor effect on μ is around **−5.58 ms** (=−2.17−3.41). This distribution, especially the effect of τ and μ on the mean difference between the two distractors, is in stark contrast with the pattern in the RRC condition.

Table 4.14: Ex-Gaussian distribution fitted to the data for the ungrammatical condition in the spillover region in Experiment 1

	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept	441.83	11.03	420.30	463.55
Distractor	-25.71	7.46	-40.50	-11.24
τ of Intercept	4.89	0.06	4.76	5.02
τ of Distractor	-0.23	0.07	-0.37	-0.09

(a) RRC condition

	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept	430.46	11.31	408.33	452.74
Distractor	-2.17	7.28	-16.54	12.11
τ of Intercept	4.75	0.07	4.60	4.89
τ of Distractor	-0.03	0.09	-0.20	0.13

(b) ARC condition

Ex-Gaussian distribution fit to data in Experiments 2 In the Experiment 2 RRC condition (Table 4.15), the mean difference between the distractors is **−22.08 ms**. The effect of τ on the mean is **−27.53 ms** (= exp(4.86+ (−0.24))−exp(4.86) = 101.49−129.02). The effect of μ on the mean is **5.45 ms** (=−22.08−(−27.53)). In the Experiment 2 ARC condition, the mean difference between the distractors is **−23.60 ms**. The effect of τ on the mean is **−31.52 ms** (= exp(4.83+ (−0.29))−exp(4.83) = 93.69−125.21). The effect of μ on the mean is **7.92 ms** (=−23.60−(− 31.52)).

Table 4.15: Ex-Gaussian distribution fitted to the data for the ungrammatical condition in the spillover region in Experiment 2

	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept	387.13	10.84	365.90	408.45
Distractor	-22.08	6.41	-34.90	-9.83
τ of Intercept	4.86	0.07	4.73	4.99
τ of Distractor	-0.24	0.07	-0.38	-0.10

(a) RRC condition

	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept	383.33	11.14	361.62	405.30
Distractor	-23.60	6.55	-36.64	-11.01
τ of Intercept	4.83	0.06	4.70	4.95
τ of Distractor	-0.29	0.08	-0.44	-0.14

(b) ARC condition

Ex-Gaussian distribution fit to data in Experiment 3 In the Experiment 3 RRC condition (Table 4.16), the mean difference between the distractors is **−11.04 ms**. The effect of τ on the mean is **−25.68 ms** (= exp(4.72+ (−0.26))−exp(4.72) = 86.49−112.17). The effect of μ on the mean is **14.64 ms** (=−11.04−(−25.68)). In the Experiment 3 ARC condition, the mean difference between the distractors is **−11.55 ms**. The effect of τ on the mean is **−20.24 ms** (= exp(4.63+ (−0.22))−exp(4.63) = 82.27−102.51). The effect of μ on the mean is **8.69 ms** (=−11.55−(−20.24)).

Table 4.16: Ex-Gaussian distribution fitted to the data for the ungrammatical condition in the spillover region in Experiment 3

	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept	372.67	10.81	351.58	394.00
Distractor	-11.04	5.54	-22.32	-0.50
τ of Intercept	4.72	0.08	4.57	4.87
τ of Distractor	-0.26	0.07	-0.41	-0.12

(a) RRC condition

	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept	371.91	10.65	351.28	392.96
Distractor	-11.55	5.33	-22.17	-1.24
τ of Intercept	4.63	0.08	4.47	4.77
τ of Distractor	-0.22	0.07	-0.36	-0.08

(b) ARC condition

Summary A summary of the findings is presented in Table 4.17, where the parameter estimates in the model outcome are converted into milliseconds. The averaged attraction effect presented in the table is the overall averaged reading time difference between the two distractor conditions (= plural − singular distractor). This effect is further decomposed into an effect of τ (distribution shift in skewness) and an effect of μ (distribution shift in mean); the averaged effect is the sum of the effects in these two parameters).

Table 4.17: The average agreement attraction effect, and the mean (μ) and skewness (τ) effects in the ungrammatical condition at the spillover region

	Exp. 1		Exp. 2		Exp. 3	
	RRC	ARC	RRC	ARC	RRC	ARC
Average attraction effect	−25.71	−2.17	−22.08	−23.60	−11.04	−11.55
τ effect on attraction	−27.31	3.41	−27.53	−31.52	−25.68	−20.24
μ effect on attraction	1.6	−5.58	5.45	7.92	14.64	8.69

Note. The attraction effect is calculated by the difference between the two distractor conditions(=plural − singular). Estimates are calculated in raw reading times (ms).

To summarize, the analyses suggest that the observed effect in our experiment cannot be solely explained by configurational differences in the position of the distractor relative to the target. This finding contrasts with observations in Staub (2010), where the relative position

of the distractor to the target led to different distributions. In our case, both intervening (Experiment 1) and non-intervening (Experiment 2) exhibited the attraction effect coming from the skewness of data points. I suspect that such a discrepancy may be due to the difference in the procedure and the nature of the dependent variable. For instance, Staub (2010) used participants' reaction times for selecting the verb (e.g., "The clubs that the advertisement..." and choosing {WAS/WERE}) after previous words were shown one word at a time; I used reading times in a self-paced reading task. Additionally, only the correct responses were used for the analysis; I used data points from incorrect (ungrammatical) verbs for the analysis. While further investigation is required to account for these discrepancies, the current additional analysis supports that the findings in our experiment are not solely due to inherent parsing differences evoked by different structures across experiments.

4.6.5 Limitations and future work

In this section, other possible factors that were not tested or not included in the experimental design are discussed. These include the type of appositive structure, statistical power and analysis, and revisiting the notion of discourse relations.

First, the current study only examined a very specific type of appositives—the non-restrictive (appositive) relative clauses. Commas were used to cue people the distinction between RRCs and ARCs. Previous research has demonstrated that readers do attend to commas despite their seemingly trivial visual salience (Angele, Gutiérrez-Cordero, Perea, & Marcet, 2023; Hirotani et al., 2006; Stine-Morrow et al., 2010; Warren, White, & Reichle, 2009). But it is possible that there is individual variability in their perception of the comma cue, and there may also be grammatical variations regarding the effectiveness of using commas to diagnose ARCs. The variations may be partly driven by factors like age and reading habits (e.g., Stine-Morrow et al., 2010), which I did not investigate in the current study. Future work on other types of appositive structures would be necessary in order to establish

broader empirical consensus. I note that there have been some promising results from a number of studies, such as Ng and Husband (2017) and McInnerney and Atkinson (2020), that revealed findings similar to the current Experiment 1, despite using a different type of appositives (*the one who/that...*).

Next concerns the statistical analysis of the interference effect. There can be arguments that a 3-way interaction of Distractor, Grammaticality, and Clause should be used for the primary analysis. However, as elaborated in the Analysis sections, our main interest was the contrast in the standard number agreement attraction effect, i.e., the 2-way interaction of Distractor and Grammaticality, between ARC and RRC conditions. This motivated us to use a nested contrast such that the 2-way interaction can be compared within the Clause factor, between the ARC and RRC condition.

The between-experiment comparison, yet, leaves concerns regarding the statistical power. When the critical ARC conditions of all three experiments were compared together, we only obtained an interaction between the attraction effect and experiments in some but not all the analyses I conducted (see Section 4.6.2). It is possible that the current sample size is not large enough to achieve sufficient statistical power for the complex interaction effect. However, an interesting possibility to consider is that the lack of power can be associated with the standard approach to analyzing the data. Most experimental work has used the averaged reading times, for example, as a measure for analysis. While mean reading times are informative, the approach of using averaged values assumes a certain distribution of the data points (e.g., Gaussian normal distribution), which may not provide a comprehensive picture of the observation. As we saw in the ex-Gaussian distribution analysis, the number agreement attraction effect was primarily driven by the data at the tail-end of the distribution, contributing to the skewness of the distribution. This suggests that our focus on mean reading times should be redirected or be accompanied by additional analysis of the overall distribution of the data points.

Additionally, the current question-based approach relies on the foundation that RRC content contributes to the main discourse whereas ARC is part of the subordinate discourse. However, as discussed in Göbel (2019), even when linguistic information is contained in an appositive structure, it can contribute to the main discourse structure if it is not *subordinate* to the preceding discourse unit but rather *coordinating*. While the ARC content in most of the experimental items used in the current study can be identified as subordinating units given their supplementary and secondary status in discourse, they have not been tightly controlled. In future work, while using the same configuration as in the current study, we can position the ARC content to be in a subordinating relation with the prior unit in one condition and in a coordinating relation in the other condition, and examine the absence or the presence of the interference effect.

4.7 Conclusion

In this Chapter, I investigated the way discourse structure information is used during the retrieval process in real-time language comprehension. The results from three self-paced reading experiments show that the retrieval interference effect, identified by the standard number agreement attraction, is constrained by discourse structure information. More specifically, I found that the impact of discourse structure cannot simply be represented in static features. Instead, the active state of discourse questions is in play. I argued and showed the way a discourse question-based account suitably explains how memory retrieval can be constrained by discourse structure information.

CHAPTER 5

RETRIEVING DISCOURSE STRUCTURE INFORMATION: PART II

In the previous Chapter, I examined the number agreement attraction effect to examine how discourse structure information is used during retrieval to resolve a subject-verb dependency. The (simplified version of the) structures used in the previous Chapter are repeated in (1). In this subject-verb agreement, the agreement controller (bold-faced) and the verb (underlined) were always in the same discourse structure. In this dependency, it is grammatically unacceptable to form the dependency with the distractor (italicized) as the agreement controller.

(1) Key examples in Chapter 4

a. Experiment 1

The waitress(,) who sat near *the girls*(,) was unhappy about all the noise.

b. Experiment 2

The musicians(,) who **the reviewer** praises so highly(,) will win a Grammy.

c. Experiment 3

Alicia met the *musicians*(,) who **the reviewer** praises so highly.

However, what if the dependency is not grammatically constrained by the type of discourse of the retrieval target? What if a dependency is formed across different discourse structures? In order to address these questions, I make use of a construction where there are two grammatically plausible antecedents (*the violinists* or *the singers*) for the pronoun (*their*), regardless of their discourse status (2). Whether the antecedents are hosted in the same (2a) or different (2b) discourse structures, both of the noun phrases are grammatically plausible candidates for retrieval. This contrasts with the case of the subject-verb

dependency.

(2) a. *The violinists* who admired *the singers* invited **their** mentors to the party.

b. *The violinists*, who admired *the singers*, invited **their** mentors to the party.

How does the parser retrieve linguistic units when the dependency is not grammatically constrained by discourse structure information? How does discourse structure information of the retrieval target (i.e., antecedents) affect pronoun resolution when they are in the same (or different) discourse structure? Would it be more difficult than when they are in the same discourse structure? Or would it be easier?

Section 5.1 provides a background on the type of linguistic construction. I then provide an overview of the experiments in this Chapter and the predictions on the results in Section 5.2. Sections 5.3–5.5 present experimental work of three self-paced reading task experiments. The findings collectively show that retrieving linguistic information from the same type of discourse structure is more costly than doing so with information from distinct discourse structures. In Section 5.6, an alternative approach to understanding the phenomenon is provided, and some limitations of the current work and suggestions for future studies are discussed. Section 5.7 concludes.

5.1 Linguistic dependency across discourse structures

As shown in (2), some linguistic dependencies require accessing linguistic content across discourse structures. Crossing the discourse boundary (or "boundary-crossing" (e.g., AnderBois et al., 2015)) is required to resolve the dependency, and examples in (3) present different empirical cases that involve such process. Bold-faced words indicate where dependencies need to be resolved; italicized words are antecedents that satisfy to resolve the dependency. Arrows indicate the direction of dependency, from the trigger to the antecedent.

(3) AnderBois et al. (2015) (pp. 97–98)

a. Presupposition

(i) [MAIN] ⇒ [SUBORDINATE]

John, who wouldn't talk to *Mary*, wouldn't talk to Susan **either**.

(ii) [SUBORDINATE] ⇒ [MAIN]

John wouldn't talk to *Mary*, who wouldn't talk to him **either**.

b. Anaphora

(i) [MAIN] ⇒ [SUBORDINATE]

Jones, who graded *each student*'s final paper, gave **them** detailed feedback.

(ii) [SUBORDINATE] ⇒ [MAIN]

Every speaker, all of **them** PhD students, gave a great talk.

c. Ellipsis

(i) [MAIN] ⇒ [SUBORDINATE]

Mary, who doesn't *help her sister*, told Jane **to**.

(ii) [SUBORDINATE] ⇒ [MAIN]

Melinda lost three *games of tennis* to Betty, who lost **six** to Jane.

The triggers and the relevant antecedents are in different clauses and discourse structures. Presupposition, anaphora, and ellipsis are resolved (only) through accessing linguistic information that is part of distinct discourse structures. The dependency chain across boundary holds in either direction, where the trigger in the main discourse structure depends on the antecedent in the subordinate structure (as in (i)), and the other way around (as in (ii)). While subordinate discourse content is part of not-at-issue content and may be represented separately from the at-issue main discourse content, the subordinate discourse content gets incrementally updated to the discourse model, and this allows the dependency across distinct discourse boundaries to be resolved (AnderBois et al., 2015).

Based on this phenomenon, I make use of the structure involving pronoun resolution where the dependency is formed either within the same or different discourse structures (4).

(4) a. [MAIN] ⇒ [MAIN] (*the violinists*) or [SUBORDINATE] (*the singers*)
The violinists, who admired *the singers*, invited **their** mentors to the party.

b. [SUBORDINATE] ⇒ [MAIN] (*the singers* or *the violinists*)
The singers admired *the violinists*, who invited **their** mentors to the party.

The key aspect of these constructions in the current work is that retrieval targets are either part of the same or different discourse structures. While the plausible antecedents are in different discourse structures in (4a), they are in the same discourse structure in (4b). But crucially, the type of discourse structure of the retrieval targets does not govern the grammaticality of the antecedent of the pronoun. To the best of our knowledge, this is the first study to examine the processing involving the boundary-crossing phenomenon.

5.2 Overview of experiments and predictions

Three sets of experiments involving inter-sentential pronoun resolution with ambiguous antecedents (as shown in (5)) were conducted using a self-paced reading task. Experiments 4–5 had 4 conditions, varied by the sentential position (sentence-medial vs. final) of the RC, and the type of RC (RRC vs. ARC). The RC was in the sentential-medial position in (i) whereas in the sentence-final position in (ii). This manipulation made the two NPs positioned either in the same discourse structure (in the RRC) or in distinct discourse structures (in the ARC condition). Experiment 5 was minimally modified from Experiment 4 to provide additional time to process the RC. Experiment 6 was a replication of the sentence-medial condition in Experiment 4 but included a modified condition that replaced *their* with *the*. This was to tease apart simply the discourse status effect from an additional retrieval interference effect.

(5) Key examples in Chapter 5

a. Experiment 4

(i) *The violinists*(,) who admired *the singers*(,) invited **their** mentors to the party.

(ii) *The singers* admired *the violinists*(,) who invited **their** mentors to the party.

b. Experiment 5

(i) *The violinists*(,) who admired *the singers* without reason(,) invited **their** mentors to the party.

(ii) *The singers* admired *the violinists* without reason(,) who invited **their** mentors to the party.[1]

c. Experiment 6

(i) *The violinists*(,) who admired *the singers* without reason(,) invited **their** mentors to the party.

(ii) *The violinists*(,) who admired *the singers* without reason(,) invited **the** mentors to the party.

If the discourse status of the retrieval targets is used for resolving the linguistic dependency, then whether the two possible antecedents are part of the same discourse or not would affect the ease of pronoun resolution. In the RRC condition in all three experiments, both antecedents are part of the same discourse structure. However, in the ARC condition, the two antecedents are part of a different discourse structure when the RC is positioned sentence-medially. When RC is in the sentence-final position, the two antecedents are part of the same discourse in the ARC condition as well. Hence, we expect to see different reading times in resolving the pronoun primarily in the sentence-medial ARC conditions.

1. This sentence is unnatural due to the added prepositional phrase but is included for the purpose of replicating the design in Experiment 4.

To preview the findings, the reading time difference is not divided by RRC vs. ARC but by the discourse structure status of the targets that need to be retrieved. Specifically, there is a greater processing cost when the retrieval targets are part of the same discourse structure. A summary of the reading time results can be found in Table 5.7 in Section 5.6.1.

5.3 Experiment 4

This experiment examines how discourse structure information is used when plausible NPs need to be retrieved for pronoun resolution. The key structure is presented in (2) (repeated below in (6)), where either *the violinists* or *the singers* are possible antecedents for *their* in both RC structures.

(6) a. The violinists who admired the singers invited their mentors to the party. [RRC]
b. The violinists, who admired the singers, invited their mentors to the party. [ARC]

The key contrast lies in the retrieval targets are located in. In (6a), both NPs are part of the main discourse structure, and at the retrieval site (*their*) to resolve the pronoun, the parser needs to access linguistic information that is hosted in the same discourse structure. In contrast, in (6b), the retrieval targets are part of discourse structures. One NP (e.g., *the violinists*) is part of the main discourse and the other (e.g., *the singers*) is part of the subordinate discourse structure information.

The empirical question is whether retrieving linguistic information from the same type of discourse structure would lead to a greater processing cost. The antecedent of the pronoun is more ambiguous in (6a) compared to (6b) since both plausible antecedents are part of the same discourse structure in (6a). Given previous findings that reference resolution becomes more costly when ambiguous pronouns are processed (i.e., ambiguity penalty) (e.g., Badecker & Straub, 2002; cf. Creemers & Meyer, 2022; Stewart, Holler, & Kidd, 2007), we can expect longer reading times to resolve the pronoun (at the *their* and the spillover regions) in (6a)

compared to (6b). On the contrary, if resolving a pronoun is easier when possible antecedents are in distinct discourse structure units, (6b) would be cognitively less burdensome and so we would see longer reading times compared to (6a).

It is possible that the contrast in (6) may simply reflect syntactic differences between the two constructions. In order to focus on the discourse-structure effect, I include a control set of conditions as well, illustrated in (7). While the contrast between RRC vs. ARC is maintained, the two plausible antecedents are both part of the main discourse structure information.

(7) a. The singers admired the violinists who invited their mentors to the party. [RRC]
b. The singers admired the violinists, who invited their mentors to the party [ARC]

If we see longer reading times at the pronoun resolution site in RRC > ARC in (6) but no difference between the two in (7), whether the antecedents are retrieved from the same discourse structure matters. However, if the RRC condition takes longer than the ARC condition in both constructions, then it is less about where the retrieval target is hosted and about the overall processing cost induced by the RRC structure.

To preview the findings, reading times at the pronoun resolution site take longer in (6a) than (6b) but no differences between (7a) and (7b). This suggests that it is the discourse structure status of the retrieval targets that affects pronoun resolution rather than simply the distinct clause type.

5.3.1 Methods

Subjects

A total of 104 native speakers of American English were recruited via Prolific. One of the subjects gave up before the end of the experiment. One participant, who reported their

first language was not English, was excluded from the main analysis. Another subject, who participated in the experiment twice, was excluded from the main analysis. This left us with 101 subjects for the analysis (mean age = 29.77 (range: 19–40); 51 females and 50 males). The duration of the experiment was approximately 15 minutes, and participants were paid $3 USD for compensation (amounting to an hourly rate of $12 USD).

Material and design

A total of 32 critical items were created with four conditions for each item (Table 5.1). They were varied by Clause type, either RRC or ARC, marked by commas. The discourse structure status of the antecedents was also manipulated. This was done so by either positioning the RC at the sentence-medial position or at the sentence-final position. The sentence-final position allowed both of the antecedents to be part of the same discourse structure. A subject-extracted relative clause was used. The sentence-medial condition had the format of "NP1(,) who VERB NP2(,) VERB *their* NP ..." and the sentence-final condition the construction of "NP1 VERB NP2(,) who VERB *their* NP ..." The NPs were all plural nouns; NP1 and NP2 were vocational nouns. The pronoun was constant as *their* across all items. The pronoun and the head noun (e.g., *their mentors*) served as the regions of interest.

The same set of items but with the two NPs swapped in their position were also included. This reversed-positioned set of items was included to examine any potential effects of placing a certain NP instead of the other at the matrix subject position in terms of the naturalness of the sentence. In addition to the critical items, 20 filler items were included in the trials. All filler trials had a pronoun, either *him* or *her*, whose antecedent was not ambiguous given the sentence (e.g., "Melanie hired Bradley, and similarly/so Malcom recruited him/her after the interview."). The filler trials were taken from the target trials in Wolf, Gibson, and Desmet (2004).

Table 5.1: An example of the experimental item for Experiment 4

RC-Position	Clause	Sentence	Discourse status of antecedents {NP 1, NP 2}
Medial	RRC	The / violinists / who / admired / the / singers / invited / **their** / **mentors** / to / the / party.	{[MAIN], [MAIN]}
Medial	ARC	The / violinists, / who / admired / the / singers, / invited / **their** / **mentors** / to / the / party.	{[MAIN], [SUBORD.]}
Final	RRC	The / singers / admired / the / violinists / who / invited / **their** / **mentors** / to / the / party.	{[MAIN], [MAIN]}
Final	ARC	The / singers / admired / the / violinists, / who / invited / **their** / **mentors** / to / the / party.	{[MAIN], [MAIN]}

Note. The '/' sign marks the region of the moving window in the self-paced reading task. Regions of interest are bold-faced. In the example *the violinists* corresponds to NP1 and *the singers* to NP2 in the sentence-medial RC condition; and the other way around in the sentence-final RC condition. Subord. = subordinate.

Procedure

A Latin-square design is used to distribute the 32 main trials and 20 filler trials. Each trial consists of a self-paced reading task, followed by an antecedent selection choice task. A word-by-word moving window paradigm (Just, Carpenter, & Woolley, 1982) is used, where the words were initially masked with dashes but later each word showed up as the participant pressed the space bar. The sentences were presented in a non-cumulative fashion, where the words were masked again with dashed lines as the participant moved on to the next word region. After reading each sentence, a forced-choice task showed up, where the participant was asked to choose the antecedent of the pronoun with a question such as "Who do you think *their mentors* refer to?" The participant was given three options, two of which were the first NP (e.g., *the violinists*) or the second NP (e.g., *the singers*), and the final option with "someone else not mentioned in the sentence." The order of the three options was ran-

domized across items and participants. No feedback was given to the participant throughout the experiment. Eight practice trials were presented before the main experiment. The implementation of the tasks including the Latin-square design was done on PCIbex Farm (Zehr & Schwarz, 2018).

Analysis

Twelve participants were excluded from the main analysis, whose accuracy with the filler trials was below 80%. This left us with 89 participants for the main analysis. The mean accuracy with the filler trials was 88.83%. Before analyzing the reading time data, I removed trials that selected the "someone else not mentioned in the sentence" as the option for the antecedent (1.31% of data). This decision was made to ensure that the participants selected either the first NP or the second NP as the possible antecedent of the pronoun. This was done to ensure that the parser was engaged in retrieving an antecedent in the given discourse. Data points above the 3 SD threshold grouped by condition and region (additional 1.46% of data removal) were also removed. The regions of interest were identified at the pronoun region (*their*) and the noun that follows (e.g., *mentors*).

Log-transformed reading time data were used for the analysis, and the analyses were done on the critical region (*their*) and the following spillover region. Linear mixed-effects regression models were built using the `lmerTest` package (Bates et al., 2015) in R (R Core Team, 2022). Clause type (RRC vs. ARC), RC-position (medial vs. final), and their interaction were included as fixed effects. In addition to these main fixed effects, log-transformed reading times from the previous region, the selection of antecedent (first NP vs. second NP), and the NP order (list A vs. list B) were also included as fixed effects to reflect spillover effects and the subjects' antecedent preference. Sum coding was used with these fixed effects.[2] The models included maximal random effects with by-subject and by-item random intercepts

2. For Clause: ARC = -0.5 and RRC = 0.5; for Position: Medial = -0.5; Final = 0.5; for NP order: List A = -0.5 and B = 0.5; for NP selection: first NP = -0.5; second NP = 0.5.

and slopes following Barr et al. (2013).[3] I used the `buildmer` package (Voeten, 2022) in R (R Core Team, 2022) to find the maximal model – random effects were simplified when the models failed to converge. I follow the convention of reaching statistical significance when the absolute t-value exceeded 2 (Gelman & Hill, 2006).

5.3.2 Results

Antecedent selection

Figure 5.1 presents the mean percent selection of choosing one of the three options as an antecedent of the pronoun. When the RC was in the sentence-medial position, the first NP was more preferred than the second NP as the antecedent, regardless of clause type. Conversely, when the RC was in the sentence-final position, there was a preference for the second NP to the first NP as the antecedent of the pronoun.

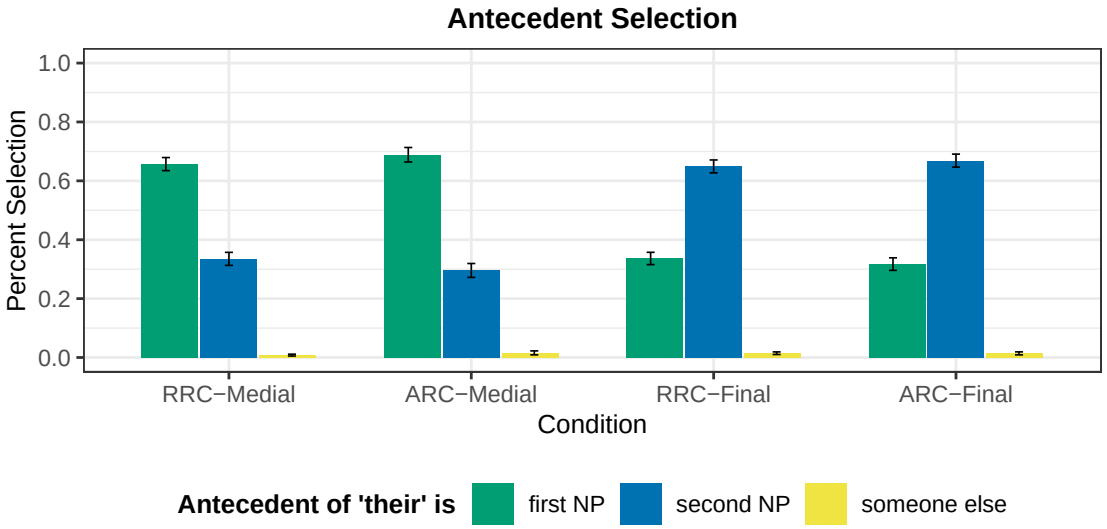


Figure 5.1: Antecedent selection task results in Experiment 4.
Note. Error bars indicate standard errors of the mean.

3. The final model for analyzing the log RT at the critical region: `logRT ~ clause*position + logRTpreviousRegion + antecedentChoice + NPorder + (1|subject) + (1|item)`. The final model for analyzing the log RT at the spillover region: `logRT ~ clause*position + logRTpreviousRegion + antecedentChoice + NPorder + (1 + clause|subject) + (1|item)`

Reading times

Figure 5.2 illustrates the mean log reading times spent at each word region (left: sentence-medial condition; right: sentence-final condition). Regression models showed that there was a statistically reliable main effect of Clause at the critical region ($\hat{\beta}$=0.053, se=0.012, t=4.569), suggesting that the RRC condition took longer than the ARC condition. There was also a main effect of Position ($\hat{\beta}$=-0.026, se=0.013, t=-2.020) in the critical region, showing that the sentence-medical RC condition took longer than the sentence-final RC condition. However, this was qualified by the 2-way interaction of Clause and Position of the RC at both the critical region ($\hat{\beta}$=-0.094, se=0.0223, t=-3.994) and the spillover region ($\hat{\beta}$=-0.089, se=0.029, t=-3.084), where the longer reading times in the sentence-medial condition mainly derived from the longer reading times in the RRC condition.

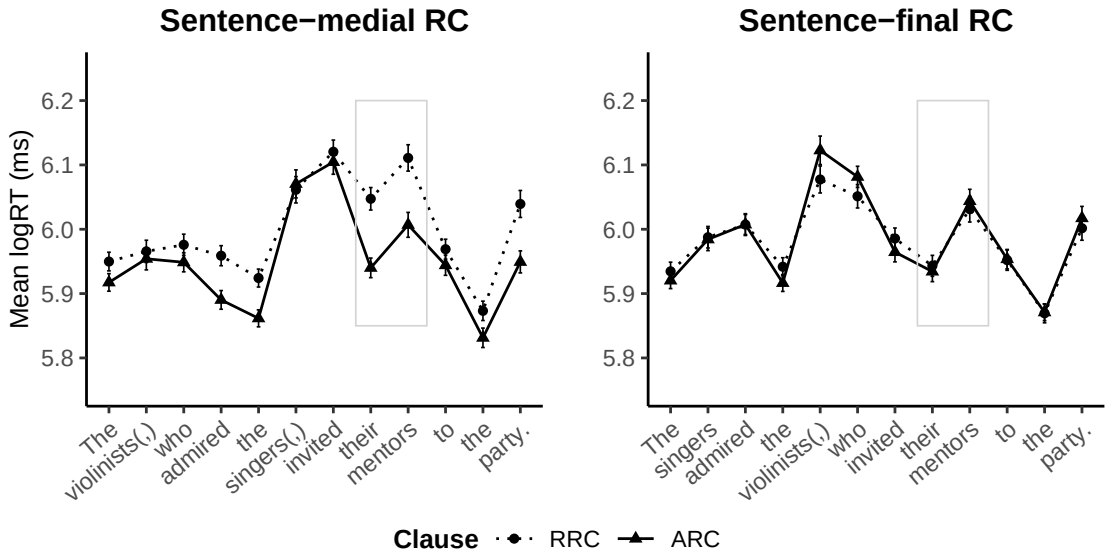


Figure 5.2: Mean log reading times in Experiment 4.
Note. Error bars indicate standard errors of the mean. Grey boxes mark the regions of interest.

5.3.3 Discussion

In the antecedent selection task, a preference for one over the other NP is found depending on the sentential position of the RC. Specifically, there was a preference for the first NP with the

Table 5.2: Summary of statistical analyses of the reading times in Experiment 4

	Estimate	Std. Error	t value
(Intercept)	4.637	0.099	46.815
Clause	**0.053**	**0.012**	**4.569**
Position	**-0.026**	**0.013**	**-2.020**
logRTpreviousRegion	0.219	0.016	13.846
Choice	0.001	0.013	0.116
List	0.002	0.012	0.203
Clause:Position	**-0.094**	**0.023**	**-3.994**

(a) Critical region

	Estimate	Std. Error	t value
(Intercept)	4.235	0.120	35.271
Clause	0.029	0.017	1.705
Position	-0.019	0.015	-1.242
logRTpreviousRegion	0.303	0.019	15.578
Choice	**0.035**	**0.016**	**2.202**
List	-0.021	0.014	-1.482
Clause:Position	**-0.089**	**0.029**	**-3.084**

(b) Spillover region

sentence-medial RC condition and for the second NP with the sentence-final RC condition. This preference is not surprising given previous work that reports a general preference for the subject NP as an antecedent for a pronominal form (J. E. Arnold, 2010, i.a.). In the sentence-medial condition, e.g., "The violinists(,) [who admired the singers(,)] invited their mentors to the party," *the violinists* (first NP) is a clausemate of the pronoun and is the subject in the corresponding clause. In the sentence-final condition, "The singers invited the violinists$_i$(,) [who$_i$ admired their mentors to the party]," *the violinists* (second NP) is structurally represented within the RC (de Vries, 2007) and serves as the subject of the clause that hosts the pronoun. In both cases, the NP that is a clausemate with the pronoun and is in the subject position of the shared clause is preferred to be selected as a more likely antecedent of the pronoun. Such antecedent preference, however, did not affect the real-time pronoun resolution, as seen in the reading time regression models.

In the reading time results, in both regions of interest, the RRC condition took longer

than the ARC condition in the sentence-medial condition but not in the sentence-final RC condition. This interaction of RC-position and clause type can be explained by the discourse structure status of the two possible antecedents between conditions. The antecedents in the RRC condition are both part of the main discourse structure whereas in the ARC condition, they are in distinct discourse units. This contrast led to the processing cost to resolve the pronoun. Such processing cost difference was not found in the control sentence-final RC condition, where the two possible antecedents were hosted in the same discourse structure. These findings suggest that rather than the structural differences between RRC vs. ARC, the discourse status of the retrieval targets affects processing difficulty in pronoun resolution.

One might wonder though whether the parser was given enough time to encode discourse structure information. Currently, it is assumed that the parser successfully encoded the distinct discourse status of the antecedents and that distinction is reflected at the retrieval site. However, the pronoun appears +1 or +2 regions away shortly after the second NP (e.g., *the singers*) was presented. Given this close proximity, it is possible that the discourse status of the second NP was not fully encoded before the retrieval site.

To address this potential confound, in the following experiment, a linguistic buffer is added after the second NP and before transitioning out of the RC boundary. This will provide the parser with sufficient time to encode the discourse structure information. We anticipate to replicate the current findings in the following experiment.

5.4 Experiment 5

This experiment is a replication of Experiment 4, with the aim of ensuring enough time is provided to the parser to encode discourse structure information. In this experiment, a prepositional phrase (PP) (e.g., *without reason*) is added after the second NP and before the RC boundary. See (8) for a comparison:

(8) Comparison of the experimental design between Experiments 4 and 5

a. Experiment 4

The violinists(,) who admired the singers(,) invited their mentors to the party.

b. Experiment 5

The violinists(,) who admired the singers **without reason**(,) invited their mentors to the party.

As in Experiment 4, we expect to see a greater processing cost to resolve the pronoun when the antecedents are part of the same discourse structure. The RRC condition, wherein the antecedents are both part of the main discourse structure, will lead to longer reading times at the pronoun resolution site compared to the ARC condition, wherein each of the two antecedents is part of the main and the subordinate discourse structure, respectively.

5.4.1 Methods

Subjects

A total of 104 American English native speakers were recruited through Prolific. One participant withdrew before completing the experiment. Three participants who identified their first language as other than English were excluded from the primary analysis. This left us with 100 subjects for analysis (average age = 30.31, range: 18–40; 46 females, 54 males). The experiment lasted approximately 15–20 minutes, and participants received $3.50 USD as compensation.

Material and design

The same 32 items and 20 filler items were used as in Experiment 4 except that Experiment 5 had a PP (e.g., *without reason*) before the end of the RC. While adopting the PP used in

Staub, Dillon, and Clifton (2017), I modified the PPs so that they attached to the verb but not the second NP—most of the PPs provide temporal information rather than locative. The experiment had a fully crossed 2x2 design as in Experiment 4, where clause type and the RC-position were manipulated. The clause had two levels, ARC vs. RRC, and the RC position was varied by the sentential position of the RC: sentence-medial vs. sentence-final. Table 5.3 illustrates a sample set of the material.

Table 5.3: An example of the experimental item for Experiment 5

RC-Position	Clause	Sentence	Discourse status of antecedents {NP 1, NP 2}
Medial	RRC	The / violinists / who / admired / the / singers / without / reason / invited / **their** / **mentors** / to / the / party.	{[MAIN], [MAIN]}
Medial	ARC	The / violinists, / who / admired / the / singers / without / reason, / invited / **their** / **mentors** / to / the / party.	{[MAIN], [SUBORD.]}
Final	RRC	The / singers / admired / the / violinists / without / reason / who / invited / **their** / **mentors** / to / the / party.	{[MAIN], [MAIN]}
Final	ARC	The / singers / admired / the / violinists / without / reason, / who / invited / **their** / **mentors** / to / the / party.	{[MAIN], [MAIN]}

Note. The '/' sign marks the region of the moving window in the self-paced reading task. Regions of interest are bold-faced. In the example *the violinists* corresponds to NP1 and *the singers* to NP2 in the sentence-medial RC condition. Subord. = subordinate.

It is worth noting that adding a PP in the RC makes the items less natural in the sentence-final condition, hence this is not a fair comparison and is not our main interest. However, we keep the sentence-final conditions for the purpose of keeping the same experimental design as in Experiment 4. The primary interest is the sentence-medial condition, examining whether the discourse status of the antecedents affects pronoun resolution, comparing the reading time at the retrieval site.

Procedure

The same procedure was used as in Experiment 4, along with the same set of items and the same type of antecedent selection question task.

Analysis

Fifteen participants were excluded from the main analysis, whose accuracy with the filler trials was below 80%. This left us with 85 participants for the main analysis. The mean accuracy with the filler trials was 89.64%. Before analyzing the reading time data, trials that selected the "someone else not mentioned in the sentence" for the antecedent (0.99% of data) were removed. Data points that were above the 3 SD threshold grouped by condition and region (additional 1.49% of data removal) were also removed. The regions of interest were identified at the pronoun region (*their*) and the noun that follows (e.g., *mentors*).

Log-transformed reading time data were used for the analysis. Given the variation in the number of words in the PP, I included the final two words in the PP region so that the final word of the embedded RC (e.g., *reason*) (including the comma for the ARC condition) and the word prior to it (e.g., *without*) were included in the PP region. Then the critical region (*their*) and the region after, the spillover region, were included in the main analysis. Linear mixed-effects regression models were built, following the same analysis method in Experiment 4.

5.4.2 Results

Antecedent selection

Figure 5.3 presents the mean percent selection of choosing one of the three options as an antecedent of the pronoun. When the RC was in the sentence-medial position, the first NP was more preferred than the second NP as the antecedent, regardless of clause type.

Conversely, when the RC was in the sentence-final position, there was a preference for the second NP to the first NP as the antecedent of the pronoun.

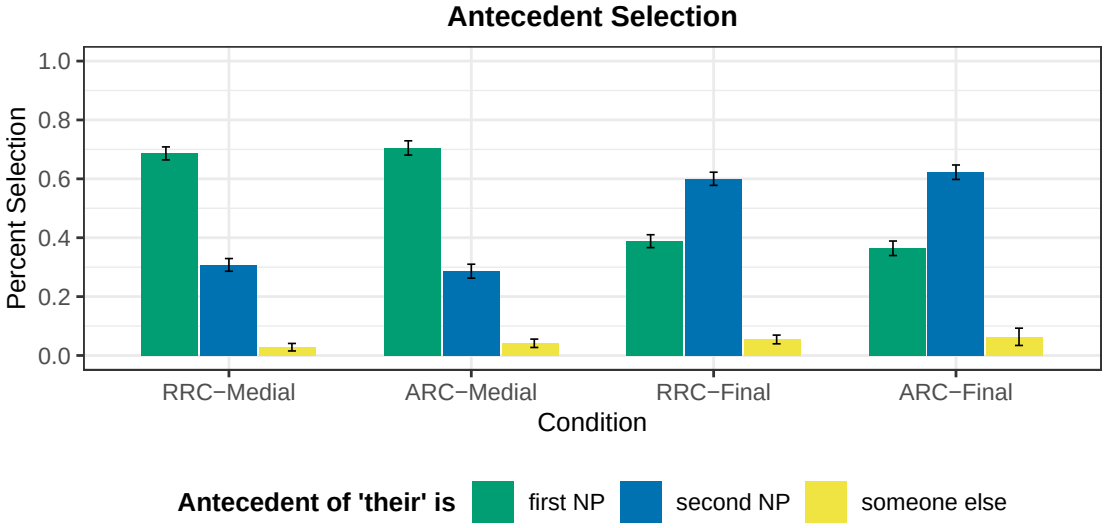


Figure 5.3: Antecedent selection task results in Experiment 5.
Note. Error bars indicate standard errors of the mean.

Reading times

Figure 5.4 illustrates the mean log reading times spent at each word region (left: sentence-medial condition; right: sentence-final condition). Regression models showed that in the critical region, there was a statistically reliable main effect of Clause ($\hat{\beta}$=0.024, se=0.012, t=2.035) but no effect of Position ($\hat{\beta}$=-0.016, se=0.013, t=-1.240), showing that the RRC condition took longer than the ARC condition regardless of the RC-position. There was also a trend of a 2-way interaction of Clause and Position of the RC ($\hat{\beta}$=0.047, se=0.024, t=1.979), indicating a pattern of RRC > ARC only in the sentence-medial and not in the sentence-final condition. In the spillover region, there was a main effect of Clause ($\hat{\beta}$=0.0.049, se=0.012, t=3.933), and a weak effect of Position, not reaching significance ($\hat{\beta}$=-0.025, se=0.013, t=-1.912). A trend of a main effect of Position ($\hat{\beta}$=-0.025, se=0.013, t=-1.192) and a 2-way interaction of Clause and Position ($\hat{\beta}$=-0.046, se=0.025, t=-1.872) were found as well.

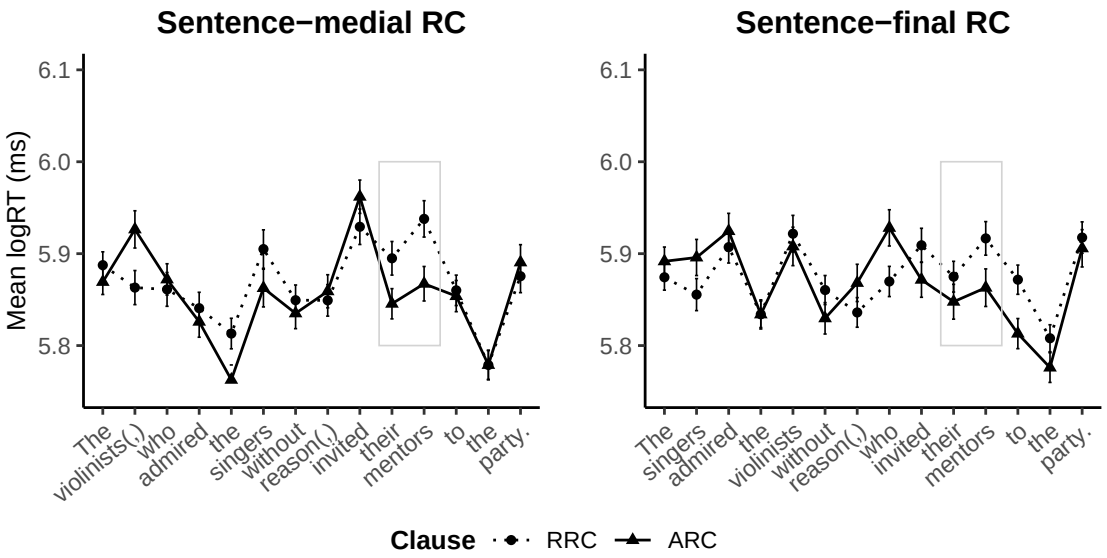


Figure 5.4: Mean log reading times in Experiment 5.
Note. Error bars indicate standard errors of the mean. Grey boxes mark the regions of interest.

5.4.3 Discussion

We replicated the findings of an antecedent selection preference observed in Experiment 4. That is, there was a preference for NP1 with sentence-medial RC conditions and NP2 with sentence-final RC conditions. This pattern held regardless of the clause type.

We found reading time differences in the sentence-final condition at regions of interest. This contrasts with the findings in Experiment 4, where no differences were found between the two clause-type conditions. It is possible that the added PP led to exhibit processing cost differences between the RRC and ARC structure. In Dillon et al. (2014), the authors found that while long-distance dependency commonly induces processing cost as more linguistic materials intervene between the dependency, this processing cost was reduced when an appositive intervened. But crucially, this reduced processing cost was evident only when the intervening materials were long enough but not when they were short. As there were longer intervening linguistic materials in Experiment 5 compared to Experiment 4, it is possible that the processing cost between RRC and ARC in the sentence-final condition was more pronounced. While this can be a plausible account, I do not pursue this further, since the

Table 5.4: Summary of statistical analyses of the reading times in Experiment 5

	Estimate	Std. Error	t value
(Intercept)	4.083	0.096	42.326
Clause	**0.024**	**0.012**	**2.035**
Position	-0.016	0.013	-1.240
logRTpreviousRegion	0.306	0.016	19.492
Choice	0.009	0.013	0.665
List	-0.018	0.012	-1.522
Clause:Position	0.047	0.024	1.979

(a) Critical region

	Estimate	Std. Error	t value
(Intercept)	3.882	0.101	38.383
Clause	**0.049**	**0.012**	**3.933**
Position	-0.025	0.013	-1.912
logRTpreviousRegion	0.338	0.016	20.479
Choice	0.006	0.014	0.455
List	0.002	0.012	0.192
Clause:Position	-0.046	0.025	-1.872

(b) Spillover region

construction used in the current experiment is not natural.

I focus on the results in the sentence-medial condition, our primary interest. The current experiment replicated the findings in Experiment 4, where the RRC condition took longer than the ARC condition at the pronoun resolution site. Since the current experimental design offered additional time before the retrieval site and after the second NP, the findings suggest that the discourse structure information of the antecedents was indeed used for pronoun resolution. This highlights the effect of the discourse status of the antecedents, particularly suggesting that retrieving pronoun referents hosted in the same discourse structure incurs processing costs.

The results in Experiments 4 and 5 collectively suggest that the discourse status of the NPs selected as pronoun antecedents is crucial in the retrieval process. A question that follows is whether the findings are a reflection of (i) a combination of the discourse status of NPs and the pronoun resolution process, or (ii) simply a reflection of the overlap of the

discourse status of the NPs. In the medial-RRC condition, both {NP1, NP2} were part of the main discourse: {[MAIN], [MAIN]}. This overlap during the encoding stage itself could have led to an interference effect, realized at the pronoun site. Hence, even when the parser was not necessarily involved in resolving the pronoun, it is possible that the condition with an overlap of {[MAIN], [MAIN]} (RRC condition) led to longer reading times at the retrieval site than the condition without the overlap {[MAIN], [SUBORD.]} (ARC condition). The following experiment aims to tease apart these two possible scenarios.

5.5 Experiment 6

The goal of this experiment is to confirm that the RRC-ARC reading time contrast demonstrated at the pronoun region in the sentence-medial condition is indeed a reflection of pronoun resolution. It is possible, as mentioned earlier, that the interference effect derived from the overlap of discourse structure information of the encoded NPs could have led to longer reading times, even without necessarily involving pronoun resolution. To test whether this is the case, I manipulate the experimental design such that the key contrast of discourse status between NPs is maintained but *their* is replaced with *the*. This manipulation is shown in (9). Only the sentence-medial RC structures are used in the current experiment.

(9) Comparison of the experimental design between Experiments 4 and 6

a. Experiment 4

The violinists(,) who admired the singers(,) invited **their** mentors to the party.

b. Experiment 6

The violinists(,) who admired the singers(,) invited **the** mentors to the party.

If the effects we observed earlier are simply due to the overlap in discourse structure status of NPs, we anticipate observing the RRC > ARC reading time difference at the pronoun

region in both *their* (replicating the result in Experiments 4) and *the*-conditions. However, if it is the case that both the discourse contrast and pronoun resolution are involved, we should expect to see the reading time difference only in the *their*-condition but not in the *the*-condition. Only when targeted entities to retrieve are present in the given discourse, i.e., the *their*-condition, we will find the reading time difference.

5.5.1 Methods

Subjects

A total of 104 American English native speakers were recruited through Prolific, but one participant withdrew before completing the experiment. Two individuals who identified their first language as other than English and one participant exceeding the specified age range were excluded from the primary analysis. Consequently, our analysis focused on 100 subjects (average age = 31.13, range: 20–40; 47 females, 52 males, and 1 non-binary individual). The experiment lasted approximately 15–20 minutes, and participants received $4 USD as compensation.

Material and design

The same set of 32 items were used as target trials and the same 20 filler items as in Experiment 4 were used, with two modifications. In the current experiment, only the sentence-medial RC structures were used, excluding the sentence-final RC configurations. Additionally, the type of determiner was included as an additional factor, featuring *their* in one condition and *the* in the other. The experiment employed a fully crossed 2x2 design, manipulating both clause and determiner. The Clause condition had two levels, ARC vs. RRC, while Determiner also featured two levels, *their* vs. *the*. Table 5.5 shows a sample set of materials. Word regions with the determiner (*their*/*the*) and the head NP (e.g., *mentors*)

were identified as the regions of interest.

Table 5.5: An example of the experimental item for Experiment 6

Determiner	Clause	Sentence
Their	RRC	The / violinists / who / admired / the / singers / invited / **their** / **mentors** / to / the / party.
Their	ARC	The / violinists, / who / admired / the / singers, / invited / **their** / **mentors** / to / the / party.
The	RRC	The / violinists / who / admired / the / singers / invited / **the** / **mentors** / to / the / party.
The	ARC	The / violinists, / who / admired / the / singers, / invited / **the** / **mentors** / to / the / party.

Note. The '/' sign marks the region of the moving window in the self-paced reading task. Regions of interest are bold-faced. In the example *the violinists* corresponds to NP1 and *the singers* to NP2 in the sentence-medial RC condition.

Procedure

A Latin-square design was used to allocate the 32 main trials and 20 filler trials. Each trial comprised a self-paced reading task, followed by a forced-choice referent selection task, following the structure of Experiment 4. The task involved questions about the referent of the pronoun (e.g., "Who do you think *their mentors* refer to?"). The participants were given three options, including the two NPs from the target sentence (*the violinists* and *the singers*) and "someone else not mentioned in the sentence." The sequence of the three options was randomized for both items and participants. Participants received no feedback during the experiment. Prior to the main experiment, eight practice trials were administered. The experiment procedure was implemented using PCIbex Farm (Zehr & Schwarz, 2018).

Analysis

The same method of data analysis was implemented as in Experiment 4. The mean accuracy score was 87.81%. Eighteen participants were excluded from the main analysis, whose

accuracy with the filler trials was below 80%. This left us with 82 participants for the main analysis. Before analyzing the reading time data, trials that selected the "someone else not mentioned in the sentence" as the option for the antecedent (9.58% of data) were removed. Additionally, data points that were above the 3 SD threshold grouped by condition and region (additional 1.58% of data removal) were removed. The regions of interest were identified at the pronoun/determiner region (*their* and *the*) and the noun that follows (e.g., *mentors*).

The analysis used log-transformed reading time data, focusing on the critical region (*their. the*) and the subsequent spillover region. Linear mixed-effects regression models were constructed using the `lmerTest` package (Bates et al., 2015) in R (R Core Team, 2022). Fixed effects included Clause type (RRC vs. ARC), Determiner (their vs. the), and their interaction. Additionally, log-transformed reading times from the preceding region, antecedent selection (first NP vs. second NP), and NP order (List A vs. List B) were incorporated as fixed effects to account for spillover effects and subjects' antecedent preferences, employing sum coding.[4] The models featured maximal random effects with by-subject and by-item random intercepts and slopes, following to Barr et al. (2013).[5] The `buildmer` package (Voeten, 2022) in R (R Core Team, 2022) was employed to establish the maximal model, with random effects simplified in cases of convergence failure. Significance was determined following the convention of an absolute t-value exceeding 2 (Gelman & Hill, 2006).

4. For Clause: ARC = -0.5 and RRC = 0.5; for Position: Medial = -0.5; Final = 0.5; for NP order: List A = -0.5 and B = 0.5; for NP selection: first NP = -0.5; second NP = 0.5.

5. The final model for analyzing log RT at the critical region: `logRT ~ clause*determiner + logRTpreviousRegion + antecedentChoice + NPorder + (1|subject) + (1|item)`. The final model for analyzing log RT at the spillover region: `logRT ~ clause*determiner + logRTpreviousRegion + antecedentChoice + NPorder + (1|subject) + (1|item)`

5.5.2 Results

Referent selection

Figure 5.5 illustrates the mean percentage selection of one of the three options as the referent. In general, the first NP was consistently chosen as the referent across conditions. In the *their*-condition, the antecedent preference observed in Experiment 4 was replicated, with NP1 selected 60% of the time and NP2 chosen 40% of the time. Conversely, when *the* used employed as the determiner, the preference for either NP was roughly equal, ranging from 40% to 50% of the selections. Furthermore, the inclination to choose someone else not mentioned in the sentence increased to 20% of the time.

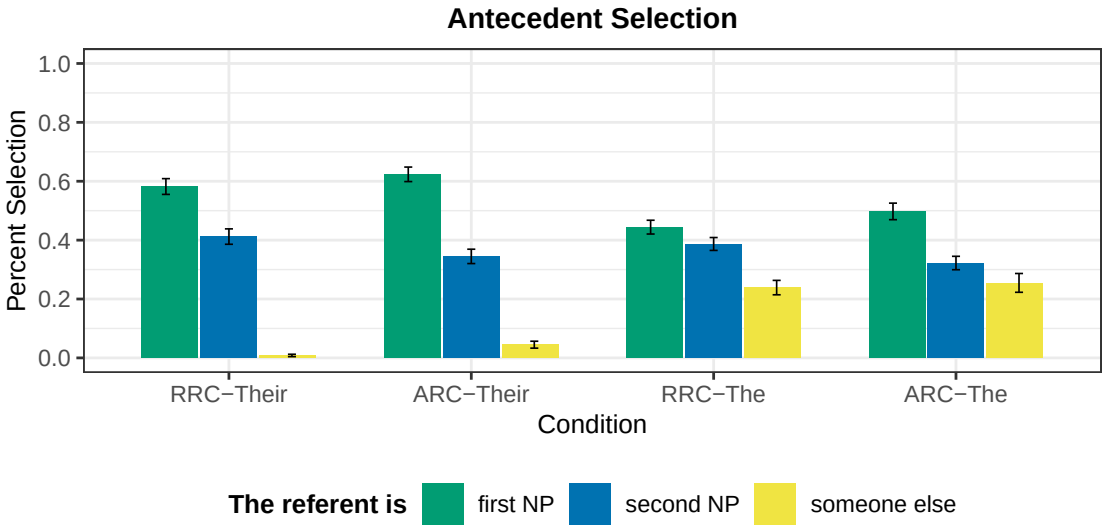


Figure 5.5: Antecedent selection task results in Experiment 6.
Note. Error bars indicate standard errors of the mean.

Reading times

Figure 5.2 illustrates the mean log reading times spent at each word region. Regression models showed that there was a statistically reliable main effect of Clause ($\hat{\beta}$=0.035, se=0.012, t=2.981) and Determiner ($\hat{\beta}$=0.050, se=0.012, t=4.331) in the critical region, showing that

the RRC condition took longer than the ARC condition, and the *their*-condition was read slower than the *the*-condition. No 2-way interaction was found ($\hat{\beta}$=-0.012, se=0.023, t=-0.537). In the spillover region, no main effect of Clause ($\hat{\beta}$=0.023, se=0.014, t=1.172) or Determiner ($\hat{\beta}$=0.022, se=0.014, t=1.593) was found. Crucially, however, a statistically reliable 2-way interaction was found ($\hat{\beta}$=0.086, se=0.027, t=3.177), indicating that the reading time difference in RRC > ARC was driven by *their*-condition but not in the *the*-condition. In other words, no clause-type difference was found in the *the*-condition and only in the *their*-condition.

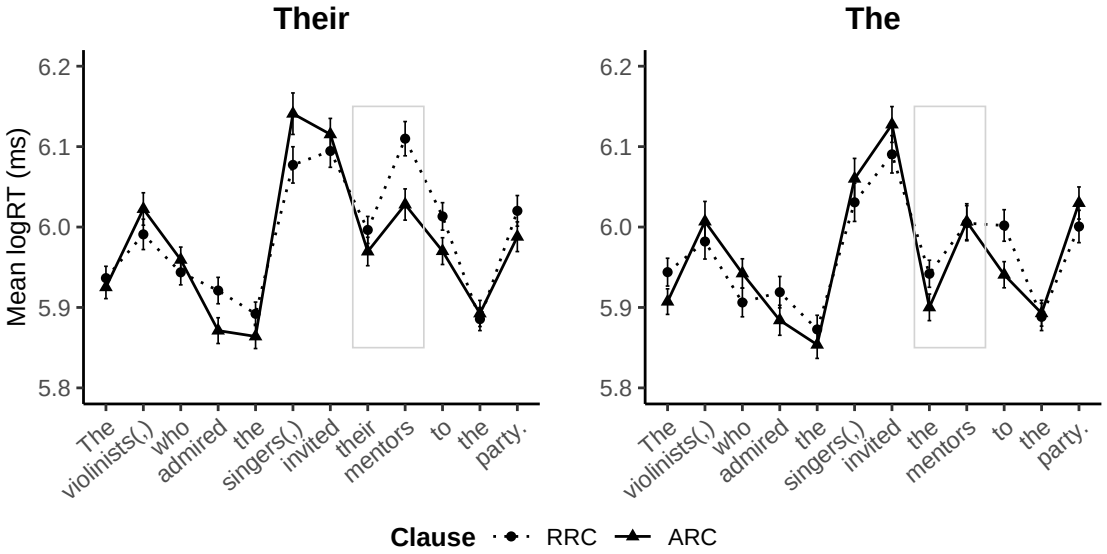


Figure 5.6: Mean log reading times in Experiment 6.
Note. Error bars indicate standard errors of the mean. Grey boxes mark the regions of interest.

5.5.3 Discussion

In the referent selection task, the division was found in the preference for choosing the referent depending on the type of the determiner. In the *their*-condition, the NP1-preference was replicated, where the first NP in the matrix clause was selected more than the second NP. This again can be interpreted as the overall preference for the subject NP for resolving pronouns (J. E. Arnold, 2010, i.a.). However, we found that the preference for choosing

Table 5.6: Summary of statistical analyses of the reading times in Experiment 6

	Estimate	Std. Error	t value
(Intercept)	5.135	0.089	57.398
Clause	**0.035**	**0.012**	**2.981**
Determiner	**0.050**	**0.012**	**4.331**
logRTpreviousRegion	0.135	0.014	9.860
Choice	0.011	0.013	0.862
List	0.011	0.012	0.936
Clause:Determiner	-0.012	0.023	-0.537

(a) Critical region

	Estimate	Std. Error	t value
(Intercept)	4.054	0.126	32.215
Clause	0.023	0.014	1.712
Determiner	0.022	0.014	1.593
logRTpreviousRegion	0.335	0.020	16.550
Choice	0.006	0.015	0.402
List	0.009	0.014	0.644
Clause:Determiner	**0.086**	**0.027**	**3.177**

(b) Spillover region

"someone else not mentioned in the sentence" as the antecedent is much higher in the *the*-condition compared to the *their*-condition. Unlike pronouns, which require a specific referent that can be uniquely identified in the discourse, the referent of "the + NP" does not need a referent available in the immediate context (e.g., B. Birner & Ward, 1994). The interpretation of an NP with a definite article is known to be less constrained to the given context. The referent can be inferred from the text without specifically being mentioned and even from the interlocutors' knowledge that has not been discussed in the discourse (e.g., Heim, 1982). This weak constraint on the interpretation of "the + NP" explains the participants' high ratio to select the "someone else" option in the *the*-condition. This selection task result also confirms that the manipulation of replacing *their* with *the* affected the parser's parsing as intended.

Another pattern to note is the selection preference in the RRC-*the* condition. This could also be explained by the general characteristics of using *the* discussed above. Unlike resolving

a pronoun, where in English there is an overall preference for a subject, anchoring back to the referent of "the + NP" has less of a constraint. It is possible that this brought the preference for the two NPs to a similar degree, amounting to 40% of selection in the current task (excluding 20% for selecting "someone else").

Next, regarding the self-paced reading task results, in the critical region, RRCs were read slower than ARCs, replicating the findings in Experiment 4. Returning to the primary goal of the current experiment—we raised the question of whether this difference of RRC > ARC will be found in both the *their* and *the*-conditions, or only in the *their*-condition. We found that the RRC-ARC difference was found in the *their*-condition but not in the *the*-condition at the critical regions.[6]

Taking together both the selection task and self-paced reading task results, the effect we see in the *their*-condition is not solely due to an effect deriving from the overlap of discourse structure status of the two NPs. Rather, it suggests that the effect is a reflection of (i) the processing cost to resolve the pronoun, and (ii) the discourse status of the antecedents used for retrieving the antecedent. The use of *their*, as opposed to *the*, requires immediate discourse and explicit antecedents to resolve the anaphoric dependency. The findings confirm that the parser was engaged in retrieving the antecedents of the pronoun in the given discourse, in the *their*-condition.

6. The RRC condition was read longer than the ARC condition in the spillover region 2 in the *the*-condition. This was not the region of analysis. But it can be interpreted to suggest that while *the* is less constrained by the immediate discourse compared to *their* (as described in the discussion about the selection task result), it does not mean it excludes any retrieval process. Retrieval for the relevant discourse entity to refer to *the mentors* could have been initiated in a delayed fashion compared to the *their*-condition.

5.6 General discussion

5.6.1 Summary of the findings

Table 5.7 provides a summary of the experimental design and the reading time results at the regions of interest. The Condition column contains the key manipulation made in each experiment. Discourse status indicates the type of discourse structure that the NPs were hosted in. This was conditioned by the experimental manipulations.

Table 5.7: Summary of the findings in Experiments 4–6

	Condition (Clause – RC-Position)	**Discourse Status** of {NP1, NP2}	**RT Finding**
Experiment 4 *their*	(a) RRC – Medial	[MAIN], [MAIN]	(a) > (b)
	(b) ARC – Medial	[MAIN], [SUBORD.]	
	(c) RRC – Final	[MAIN], [MAIN]	(c) ≈ (d)
	(d) ARC – Medial	[MAIN], [MAIN]	
	Condition (Clause – RC-Position)	**Discourse Status** of {NP1, NP2}	**RT Finding**
Experiment 5 *their* with PP in RC	(a) RRC – Medial	[MAIN], [MAIN]	(a) > (b)
	(b) ARC – Medial	[MAIN], [SUBORD.]	
	(c) RRC – Final	[MAIN], [MAIN]	*(c) > (d)
	(d) ARC – Medial	[MAIN], [MAIN]	
	Condition (Clause – Determiner)	**Discourse Status** of {NP1, NP2}	**RT Finding**
Experiment 6 *their* vs. *their* sentence-medial RC	(a) RRC – *their*	[MAIN], [MAIN]	(a) > (b)
	(b) ARC – *their*	[MAIN], [SUBORD.]	
	(c) RRC – *the*	[MAIN], [MAIN]	(c) ≈ (d)
	(d) ARC – *the*	[MAIN], [SUBORD.]	

Note. Subord. = Subordinate. RT Finding = Reading time comparison between conditions in the regions of interest. * = Condition that contains unnatural sentential structure.

One of the key findings is the constant longer reading times in (a) compared to (b) conditions across experiments. This shows that the discourse structure information is used

in retrieval, affecting processing difficulty. Specifically, retrieving the target from the same type of discourse structure leads to processing costs. We also found that the increased processing cost is not simply due to an overlap of discourse structure information where retrieval targets are hosted. If this were the case, we would have seen longer reading times in (c) > (d) as well as (a) > (b) in Experiment 6; however, this was not the case. Hence, the findings collectively suggest that the parser was engaged in pronoun resolution, i.e., retrieval process, and discourse structure information was used for retrieving the antecedents.

5.6.2 An alternative approach

Information-flow account

Instead of discourse structure, one can take an "information-flow" (Gibson, Desmet, Grodner, Watson, & Ko, 2005) perspective to compare the processing of RRC and ARC structures. This approach is based on the view that there is a more natural way of building and developing information in an utterance. Specifically, it relies on the idea that it is more natural for an utterance or a sentence to flow from old, background information to new information, appearing later. In terms of the information status of ARC, Gibson et al. suggest that ARC (or what the authors referred to as "non-restrictive relative clause" content) is usually discourse-new whereas RRC content is given information in discourse context. Hence, the prediction is that putting RRC (discourse old) at the beginning of the sentence is more natural than doing so with ARC (discourse-new) content. Based on their information-flow account, sentence-medial RRC content will be easier to process than a sentence-medial ARC condition. For example, RC content will be read faster in (10a) than in (10b). They found empirical evidence supporting this prediction.

(10) Sentence-medial RC conditions in Experiment 2 in Gibson et al. (2005)

a. The director that the critics praised at the banquet insulted an actor from a

big action movie during an interview. [RRC]

b. The director, who the critics praised at the banquet, insulted an actor from a big action movie during an interview. [ARC]

However, this approach does not correctly account for the findings in the current experiments. With regards to the reading times between sentence-medial RRC and ARC conditions, we found that the ARC condition took a shorter reading time than the RRC condition (Experiments 4 and 5). This is in stark contrast to the prediction made in Gibson et al. (2005). Since the regions of interest were after the RC in the current experiments (for the sentence-medial conditions) but within RC in Gibson et al. (2005), one might argue that it may not be a fair comparison. However, when the reading times post-RC regions in Gibson et al. (2005) (reported in their Appendix section) were compared, the pattern was still the same, where sentence-medial RRC $<$ sentence-medial ARC. Again, this is the opposite direction of the reading time results. The information-based account does not explain the current findings.

Yet, this is not to conclude that the information-flow account does not hold. It is possible that this opposite direction of reading time results could be a reflection of other factors such as the type of embedded RC. In Gibson et al. (2005), an object-extracted RC was used across conditions (as exemplified in (10)), while in the current experiment, a subject-extracted RC structure was used. The different types of extraction involved in the embedded RC structure leading to different processing costs have been widely reported in the literature (e.g., Gibson, 1998, 2000; Hale, 2001; Levy, 2008; cf. Lowder & Gordon, 2021; Staub et al., 2017). It is thus possible that object-extracted vs. subject-extracted RCs involving RRC and ARC structures could have led to the opposite processing configuration. This remains as future work to investigate. However, for understanding the empirical finding in the current experiments, where the subject-extracted RC structure was used, the discourse structure-based approach offers a more accurate explanation than the information flow-based account.

5.6.3 Limitations and future work

Discourse status of the pronoun

A potential confound in the key construction used in the current work can derive from the position of the pronoun. I primarily focused on the construction where the pronoun was positioned in the matrix clause, with the relative clause placed sentence-medially. Experiment 4, however, while manipulating the discourse structure status of the two possible antecedents of the pronoun, the discourse status of the pronoun varied. In the RC-sentence-medial condition, e.g., *The singers, who admired the violinists, invited their mentors to the party*, the pronoun was part of the main discourse; in contrast, in the RC-sentence-final condition, e.g., *The violinists admired the singers, who invited their mentors to the party*, the pronoun was part of the subordinate discourse. As pronoun resolution requires access to syntactic, semantic, and discourse-related information, the discourse status of a pronoun can be an influential factor.

Furthermore, the position of the relative clause itself can also be a confound. As Syrett and Koev (2015) find from experimental evidence, sentence-final appositives tend to behave *at-issue*-like, being part of the main discourse. Similar arguments can be found in AnderBois et al. (2015); Göbel (2019) and Wilke (2023) show through a series of experimental work regarding the effect of the sentential position of ARCs (appositives). This means that it could be the case that the sentence-final RC conditions compared in Experiment 4—*The violinists admired the singers who invited their mentors to the party* (RRC) and *The violinists admired the singers, who invited their mentors to the party* (ARC)—in practice had the same *flavor* of discourse status. If this is the case, interpreting the results of the sentence-final condition could be influenced not (just) by the type of discourse status of the antecedents but (also) by the uniqueness of a sentence-final RC.

Given these potential confounds, in future work, I aim to control the position of the

pronoun while maintaining the contrast in discourse structure information between the two antecedents. The examples in (11) illustrate a possible set of conditions that can be used.[7] The pronoun is bold-faced, and the two candidates for the antecedent, which is ambiguous, are italicized.

(11) Conditions for future work investigating encoding of discourse structure information in pronoun resolution in boundary-crossing phenomenon

a. *The singers*(,) who *the violinists* introduced to the conductor(,) invited **their** mentors to the party.

b. The singer(,) who *the violinists* introduced to *the conductors*(,) invited **their** mentors to the party.

c. The singer introduced *the violinists* to *the conductors*(,) who invited **their** mentors to the party.

The first condition (11a) is similar to the experimental design in the current study, where in the ARC condition (with commas), the two possible antecedents are part of the different discourse structure: {*the singers*, *the violinists*} = {[MAIN], [SUBORD.]}. This can be compared with the second condition (11b), where in the ARC condition, the two possible antecedents are part of the same in discourse structure: {*the violinists*, *the conductors*} = {[SUBORD.], [SUBORD.]}. The comparison of the first two conditions allows us to examine the effect of discourse structure overlap between retrieval targets. By adding the third condition (11c), we can further examine whether it is simply the discourse structure information overlap or whether the type of discourse structure matters. The third condition (11c), similar to the second condition (11b), also exhibits discourse structure information overlap between the two antecedents, but both are part of the main discourse: {*the violinists*, *the singers*} = {[MAIN], [MAIN]}. This set of conditions will allow us to investigate the same question but

7. I thank Andrew McInnerney for his suggestion on this direction and material design.

more specifically target the influence of discourse structure information overlap, avoiding the confound of the position of the pronoun.

5.7 Conclusion

In this Chapter, I examined the way the parser retrieves targets that are part of the same or different discourse structures. This was done so by making use of intra-sentential anaphoric dependency involving pronoun resolution. Two grammatically plausible antecedents of the pronoun were given within the same sentence. But they were varied by the type of discourse structure that they were hosted in. This led to an experimental design where in certain cases, the two NPs were in the same discourse (i.e., main) whereas in the other in a distinct discourse structure (i.e., NP1 in main and NP2 in subordinate discourse). The results from three self-paced reading task experiments, along with the antecedent selection task, suggested that discourse structure information is used for resolving a pronoun, even when it is not necessary information for resolving the dependency. They showed that retrieving antecedents is cognitively more costly when they are part of the same discourse structure.

CHAPTER 6

ENCODING DISCOURSE STRUCTURE INFORMATION

Findings from Chapters 4–5 suggested that discourse structure information is used for retrieval. We also found in Chapter 5 that retrieving targets that are part of the same discourse structure is more costly than those in distinct discourse structures.

In this Chapter, I further explore how discourse structure information is used during ENCODING. I examine the way the distinction in discourse structure (main vs. subordinate) affects real-time language comprehension. Specifically, I investigate whether the overlap in discourse structure information between linguistic representations during encoding leads to a greater processing cost. I also address the question of time, examining when the overlap effect during encoding is realized.

To address these questions, the same structure used in the previous Chapter will be used. The key conditions are again presented below:

(1) a. The violinists who admired the singers invited their mentors to the party. [RRC]
b. The violinists, who admired the singers, invited their mentors to the party. [ARC]

As in the previous Chapter, this construction is useful for the current purpose since discourse structure information is not necessarily a retrieval cue but is used during encoding. While the distinction of RRC (main discourse) vs. ARC (subordinate discourse) is made, discourse structure information is not necessarily for resolving the pronoun because the noun phrase in either main or subordinate discourse structure can be the antecedent. The primary focus is the degree of overlap in discourse structure information between the two possible antecedents (e.g., *the violinists*, *the singers*). Crucially, the conditions differ in the degree of the overlap in discourse structure information between the two NPs (*the singers*, *the violinists*): overlap in (1a) {[MAIN], [MAIN]} but no overlap in (1b) {[MAIN], [SUBORDINATE]}.

Previous studies have shown that overlap in linguistic representations during encoding

leads to interference, either during the encoding stage or when retrieval is triggered (e.g., Hofmeister & Vasishth, 2014; Sekerina et al., 2016; Villata et al., 2018). Interference effect during encoding can lead to competition for features between entities, which can result in degraded/distorted representations in memory (Nairne, 1990, 2002; Oberauer & Kliegl, 2006; Oberauer & Lange, 2008). The competition effect due to overlap in linguistic representation is typically realized as longer reading times (e.g., Gordon et al., 2001) or fewer eye-gazes on possible targets (e.g., Sekerina et al., 2016).

Based on these existing findings, I examine whether the overlap in discourse structure information during encoding leads to a competition effect, and investigate when that effect is realized. To this end, eyetracking experiments using a visual world paradigm are conducted. Through two web-based eyetracking experiments, (i) eye-gaze towards targets, specifically the eye-gaze differences between the targets, and (ii) the time window when the eye-gaze differences are meaningfully different between the two conditions will be analyzed.

This Chapter begins with an overview of the visual world paradigm (VWP) (Section 6.1). Comprehension experiments that use the VWP paradigm, including those that investigated pronoun resolution and competition effect are also summarized (Section 6.1.3). Section 6.3 provides an overview and predictions of the experiments. Sections 6.4–6.5 present two web-based visual world paradigm eyetracking experiments. The findings collectively show that discourse structure information is used during the encoding stage, and the overlap in the information leads to a competition effect between the encoded linguistic items. The effect can be realized as early as right after the encoding, crucially, even before retrieval is triggered. Section 6.6 addresses some limitations of the experiments and discusses future directions. Finally, I conclude in Section 6.7.

6.1 Visual world paradigm (VWP)

This section provides some general background on the VWP using an eyetracking method. Since I use a web-based eyetracking method, I also discuss the differences between the offline and a web-based method and the key setup in the current experiments.

6.1.1 Core background

The VWP involves experimental methods of monitoring eye movements on objects or pictures on a visual environment or screen as participants listen to or produce language (see Huettig, Rommers, & Meyer, 2011; Salverda & Tanenhaus, 2018, for an overview). This paradigm, as coined by Tanenhaus and colleagues (Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), is grounded on the idea that the presented visual workspace predefines a context in which language takes place. The linking hypothesis behind this paradigm is that saccadic eye movement takes place in tight association with speech and language processing (Cooper, 1974), causing shifts in visual attention towards an object as a result of planning or comprehending an utterance. Consequently, these eye movements offer insights into real-time language processing. Cooper's seminal work showed findings regarding eye movements on pictures on a display while participants were listening to stories. Participants tended to look at objects in the visual world that were named or those that were associated with those names. For instance, they rapidly shift their gaze towards elements closely associated with the meaning of the language being heard, such as looking at a lion upon hearing the word, *lion*, or looking at a lion or zebra upon hearing the word, *Africa*, even as words are being pronounced.

The type of displayed stimuli can be classified into three groups (see Huettig et al., 2011, for an overview). A popular type uses line drawings of objects with having them displayed in a predefined visual space on a computer monitor screen (Allopenna et al., 1998; Tanenhaus et al., 1995, a.o). Typically the monitor screen is designed as a visual environment in a

quadrant space, or n-by-n space, wherein objects are presented in arrays in a designated area. This type of design is best suited when investigating the process of activation of lexical and conceptual representations. Another type involves a visual scene, where objects in line drawings are presented in a semi-realistic setting (Altmann & Kamide, 1999, a.o). Since surrounding visual context in addition to objects is provided, this type of presentation is used when research questions are related to investigating participants' incorporation of world knowledge and its effect on language processing. Finally, printed words are also used as visual stimuli (McQueen & Viebahn, 2007, a.o). This type of presentation allows greater flexibility in the range of words that can be used including non-concrete, abstract ones. Research has shown that this format is better suited for studies on orthographic perception (Salverda & Tanenhaus, 2010) and research involving phonological manipulations (Huettig & McQueen, 2007) compared to exploring semantic representations.

Given the display of visual stimuli, the primary focus is usually on examining eye movements triggered by the auditory stimuli (i.e., linguistic input) and the relevant linguistic information. Researchers are typically interested in examining the eye movements within the temporal windows that are time-locked to the relevant linguistic information. Visual stimuli are commonly identified as *target*, *competitor*, and *distractor*, depending on the experimental design. The picture of interest that is the most relevant target corresponding to the unfolding linguistic input is referred to as the *target*. Potentially relevant but not the target visual object is the *competitor*. Those that are neither the target nor the competitor are *distractors*. The main interest lies in identifying the time point at which the eye gaze towards the target differs from that towards the non-targets. It has been reported that eye fixation on relevant visual objects typically occurs within 200 ms after word onset or at the *point of disambiguation* (Matin, Shao, & Boff, 1993; Salverda, Kleinschmidt, & Tanenhaus, 2014; Saslow, 1967), i.e., the time at which the entity associated with the speech stream is identified through the integration of information from linguistic input. The most popular

method of analyzing eye gaze is examining the proportion of eye fixations of the picture candidates (e.g., target, competitor, distractor) relative to the relevant time point, e.g., the onset of (or around 200 ms prior to) the critical word. The proportion of fixations, their differences between picture candidates, or the speed of divergence in eye fixations are possible measure variables in VWP experiments.

Along with the tracking of eye gazes, a task or action is also required to be conducted by the participants. participants are also assigned a task or action. They are instructed to engage in activities related to the displayed objects on the screen, such as relocating an object or clicking on it when mentioned in the auditory stimuli. These motor tasks are accompanied by instructions presented as auditory input, while simultaneously eye gazes are tracked. In different cases, no active engagement is required; instead, participants are directed to listen to the auditory stimuli and observe the objects on the screen. In such instances, participants' attention is directed to the objects mentioned in the auditory input.

6.1.2 Web-based method

More recently, the method has been used in a web-based setting as well. Papoutsaki et al. (2016) developed a web-based eye-tracking library, known as `WebGazer.js`. Based on this method, a number of researchers have used this JavaScript-based toolkit to conduct VWP (Degen, Kursat, & Leigh, 2021; Semmelmann & Weigelt, 2018; Slim & Hartsuiker, 2022; Vos, Minor, & Ramchand, 2022) even in different disciplines of study (e.g., Madsen, Júlio, Gucik, Steinberg, & Parra, 2021; Yang & Krajbich, 2021).

Web-based VWP has the same linking hypothesis as a lab-based VWP. Yet, the way eye gazes are measured (i.e., estimated and calculated) are different. Furthermore, given that data are collected through the participants' personal web cameras and through the internet, the two methods have different spatiotemporal resolutions. Here I compare the lab-based and web-based VWP methods in terms of their (i) apparatus, (ii) temporal resolution, and

(iii) spatial resolution.[1] Practical measures to address the low spatiotemporal resolution of the web-based method are also discussed.

Apparatus

In conventional lab-based eyetracking experiments, the most common apparatus used are EyeTribe, SMI, SR Research (EyeLink), and Tobii, to name a few (see Niehorster, Cornelissen, Holmqvist, Hooge, and Hessels (2018) for a review).

For using a web-based eyetracking method, a number of libraries/packages have been used, initially for crowdsourcing and online game purposes. These include TurkerGaze (Xu et al., 2015), or GazeParser (Sogo, 2013) (see Vos et al. (2022) for an overview on other toolkits using eye tracking technologies online). Others with more of a purpose of scientific research include WebGazer.js (Papoutsaki et al., 2016), and other javascript-based platforms such as jsPsych (de Leeuw, Gilbert, & Luchterhandt, 2023), or Gorilla (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020), and PsychoPy (Peirce et al., 2019). PCIbex (Zehr & Schwarz, 2018) also provides an eyetracking module, which integrates the WebGazer.js API. In the current study, I conduct web-based VWP eyetracking experiments using PCIbex with the `EyeTracker` element.

Spatial resolution

Spatial resolution is generally identified by the accuracy and precision of the eyetracker. Accuracy is calculated by the difference (i.e., offset) between the true gaze position and the recorded position. Precision is defined by the dispersion of the recorded eye gazes.

A lab-based eyetracking method tracks eye gaze and movements through invisible infrared

1. When comparing the quality of eye tracking data, commonly four of the following components are measured: spatial accuracy, spatial precision, temporal accuracy, and robustness (e.g., data loss/recording ratio) (Holmqvist et al., 2011). In this section, spatial resolution roughly covers the components of spatial accuracy and precision. The issue of robustness is covered in discussing temporal resolution.

illuminated on the participants' eyes. The camera picks up the images of the eyes, and the position and movements of the eyes are calculated based on the reflected infrared information. Accuracy is usually 0.4° to 0.9° (e.g., Tobii).

Web-based eyetrackers locate the position of the pupil center through the participants' personal web camera. No video recording is required (unless facial expressions and movements are of interest for research). Instead, the x, y coordinates of the eye gaze on the screen are recorded along with the time (t) information ([x, y, t]). Accuracy is usually 4° (Semmelmann & Weigelt, 2018) to 5° (Vos et al., 2022).

In the current experimental setup, to ensure a certain threshold of high spatial resolution, we include a calibration stage. Before the main trials and between trials, participants were asked to stare at a visual target that appeared on a screen. The Euclidean distance between the estimated eye gaze and the true target position was calculated, and the participants were able to proceed only when this distance was above a certain threshold. More specific details of the calibration procedure used in the current experiments are presented in Section 6.4.1.

Temporal resolution

The temporal resolution of an eyetracker is commonly indexed by sampling rate (or sampling frequency), which shows the number of samples that are collected/recorded each second. For example, an eyetracker with a sampling rate of 300 Hz means data is recorded every 3 ms; a sampling rate of 1000 Hz means there are 1000 samples per second (= 1000 ms).

In lab-based settings, sampling rate ranges widely differ depending on the apparatus (e.g., 30 Hz, 60 Hz, 120 Hz, ..., 1000 Hz). Language-related studies commonly use apparatus with a sampling rate of 120 Hz or 60 Hz (Dalmaijer, 2014; Ooms, Dupont, Lapon, & Popelka, 2015).

The temporal resolution of web-based eyetrackers also varies. Previous studies have reported a range of 20.73 Hz (Vos et al., 2022) to 40.2 Hz (Yang & Krajbich, 2021), and

in the current experiment the rate is 47.71 Hz (Experiment 7). The variation seems to be driven primarily by the camera latency across participants (Semmelmann & Weigelt, 2018), and their web browser performance (Vos et al., 2022).

Practical measures

As discussed in Vos et al. (2022), due to the low spatiotemporal resolution of the web-based eyetracking method, using it for types of research that require fine-grained gaze information (e.g., eyetracking-while-reading) is not recommended yet. However, the authors present a replication of a lab-based visual world paradigm eyetracking study, showing that results and effects (and effect sizes) in a lab setting can be obtained through a web-based method. Other work has also shown the validity of conducting eyetracking experiments using a visual world paradigm that presented 2–4 entities on the screen (e.g., Degen et al., 2021; Lee & Kaiser, 2022; Slim & Hartsuiker, 2022; Vos et al., 2022). Per these previous studies, I also use a visual world paradigm that contains 4 images on the screen.

To handle the low spatiotemporal resolution of the web-based eyetracking method, prior work has used calibration scores or the users' internet speed to filter poor data points (e.g., Slim & Hartsuiker, 2022). The most popular method involves making use of *calibration*. Calibrating the participants' true eye gaze and the estimated eye gaze is crucial for a correct measurement, and it is a standard procedure in a lab-based setting. Slim and Hartsuiker (2022), in their web-based study, showed that calibration score is correlated with temporal and spatial resolution, and their replication studies showed data with higher resolution are obtained when the calibration threshold is increased from 5 to 50. Recent work has set the calibration threshold score at 50 (Degen et al., 2021; Slim & Hartsuiker, 2022; Vos et al., 2022) or higher at 60 (Lee & Kaiser, 2022). This score can be lower for between-trial calibration stages, where some studies choose not to have a threshold (Lee & Kaiser, 2022) (i.e., setting the score to 0) and instead use a post-experiment filtering process. In other

cases, the frequency of calibration during trials can be increased. For example, the calibration can be done after every trial (Lee & Kaiser, 2022; Slim & Hartsuiker, 2022), or after every 10 (Yang & Krajbich, 2021) or 12 (Vos et al., 2022) trials, or every block (Semmelmann & Weigelt, 2018).

Given the comparatively recent application of this method to language research, there is not yet a rule of thumb on the threshold of calibration score or the frequency of the calibration. However, Vos et al. (2022) showed that the calibration score incrementally drops over trials, highlighting the importance of between-trial calibration procedures for higher gaze accuracy. Following their approach, I include the calibration procedure between trials as well as at the beginning of the experiment in the current experimental setup. Hence, there are two types of calibration processes. In Experiment 7, both the initial and between-trial calibration threshold scores are set high (e.g., 60 and 50 out of 100). In Experiment 8, these scores are set lower (e.g., 30) but I apply the same high threshold during the analysis.

6.1.3 Visual world paradigm in language comprehension research

Using the temporal information gained through eye gaze, the VWP has been employed for investigating word recognition, phonemic processing, syntactic parsing, predictive processing, semantic and pragmatic processing, sentence production, and reference resolution.

Regarding syntactic parsing, for instance, much scholarship has examined eye gazes and saccades in auditory stimuli containing syntactic attachment ambiguity and found evidence for the time point at which contextual information is incorporated and used for syntactic parsing (Chambers, Tanenhaus, & Magnuson, 2004; Snedeker & Trueswell, 2004; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus et al., 1995). Studies have also found evidence for predictive processing during real-time language comprehension through early saccades and eye movements to visual objects (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003). More looks at the relevant target were found even before the

onset of the corresponding word is unfolded. The results suggested that thematic information conveyed by prior words (e.g., verb before an object noun was presented) was used to evaluate the fitness of a possibly upcoming word, indicating an active predictive process in language comprehension. Other work has investigated lexical processing such as finding that semantic overlap between visual information (Huettig & Altmann, 2005), or grammatical gender information (Dahan, Swingley, Tanenhaus, & Magnuson, 2000) affects language processing at an early stage, language processing when given early phonological information (Allopenna et al., 1998; Huettig & McQueen, 2007; Salverda et al., 2014). Semantic and pragmatic inference involving implicatures, for example, has also been examined, investigating when such computation occurs (Degen & Tanenhaus, 2016; Y. T. Huang & Snedeker, 2009; Y. T. Huang, Spelke, & Snedeker, 2013; Snedeker & Trueswell, 2004; Sun & Breheny, 2020). The paradigm has also been proven suitable for examining language processing among pre-literate children and language development (Y. T. Huang & Snedeker, 2009; Y. T. Huang et al., 2013; Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill, & Logrip, 1999), population with language disorder (Thompson & Choy, 2009) or neurodivergence (Brock, Norbury, Einav, & Nation, 2008), and for sentence production studies (Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998).

Work has used the VWP to investigate reference resolution, involving anaphors (J. E. Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Clackson, Felser, & Clahsen, 2011; Cozijn, Commandeur, Vonk, & Noordman, 2011; Han, Moulton, Block, Gendron, & Nederveen, 2021; Kaiser, Runner, Sussman, & Tanenhaus, 2009; Runner & Head, 2014; Runner, Sussman, & Tanenhaus, 2003, 2006), reflexive pronouns (Clackson et al., 2011; Clackson & Heyer, 2014; Han et al., 2021; Kaiser et al., 2009; Runner & Head, 2014; Runner et al., 2003, 2006), and demonstratives (Brown-Schmidt, Byron, & Tanenhaus, 2005; Kaiser & Trueswell, 2008). These studies have shown different types of information that are in play for reference resolution such as salience, semantic and syntactic structures, and gender information.

Several studies have used the paradigm to examine real-time pronoun resolution, addressing the types of information involved in the resolution and the timeline at which access to such information affects parsing. For example, Clackson et al. (2011) examined a case in which two antecedents could potentially compete due to gender feature overlap. An example from Clackson et al. (2011) Experiment 2 is given in (2).[2]

(2) Pronoun condition in Experiment 2 in Clackson et al. (2011)

a. Gender feature overlap

Peter was waiting outside the corner shop. *He* watched as Mr. Jones bought a huge box of popcorn for **him** over the counter.

b. No gender feature overlap

Susan was waiting outside the corner shop. *She* watched as Mr. Jones bought a huge box of popcorn for **her** over the counter.

Due to binding constraints, there was only one grammatical antecedent for the bold-faced pronouns in (2b) and (2a), namely the matrix subject. The intervening noun phrase, Mr. Jones, either overlaps (2a) or does not overlap (2b) in gender with the grammatical antecedent.Four images were presented on the screen as each stimulus sentence was auditorily presented. The images included the two characters (e.g., Susan and Mr. Jones), the object (e.g., *popcorn*) mentioned in the stimulus sentence, and one distractor inanimate object not mentioned in the stimulus. More looks were found towards the grammatical antecedent (e.g., looks towards Susan upon hearing *her* in the (2b) condition). Crucially, however, the proportion of looks to the target was reduced in the overlap condition (2a), suggesting the overlap in gender features between the grammatically licit and illicit entities leads to "competition," which was manifested as a smaller eye gaze difference between the two relevant

2. Both pronouns and reflexive pronouns, and children and adult speakers were compared, but here I present results from the pronoun condition among adult speakers.

images. Similar observations were also reported in Runner and Head (2014).

Building upon Runner et al. (2003) and Runner et al. (2006), Kaiser et al. (2009) used a VWP to investigate the processing of the so-called picture noun phrase. They investigated the role of structural and discourse-pragmatic factors influencing (reflexive) pronoun resolution, using auditory stimuli with structures as in (3). Reflexive pronouns were compared but I present the example of the pronoun condition only.

(3) Pronoun condition in Experiment 3 in Kaiser et al. (2009)

a. Subject: source of information; Object: perceiver
Peter told *Andrew* about the picture of **him** on the wall.

b. Subject: perceiver; Object: source of information
Peter heard from *Andrew* about the picture of **him** on the wall.

Two characters mentioned in the stimuli were displayed on the screen; for example, two male characters were presented, and below each had a name label, Peter and Andrew. In this so-called picture noun phrase structure, either the subject of the matrix clause or the embedded subject is a possible referent of the pronoun. Given such ambiguity, the authors found that more looks fell onto the perceiver (e.g., Andrew in (3a) and Peter in (3b)). Since looks primarily fell onto the matrix subject when a reflexive pronoun was used, the authors suggested that discourse-pragmatic information is readily used when resolving pronouns.

6.2 Time window of competition effect

As earlier studies show, overlap in linguistic representations during the encoding stage can lead to an interference effect (see Section 2.2 for a detailed description of the interference effect during encoding). The feature similarities between linguistic representations during encoding can result in competition for the shared features. The competition can lead to a

degraded representation in memory (Nairne, 1990, 2002; Oberauer & Kliegl, 2006; Oberauer & Lange, 2008).

Empirical findings diverge as to when the interference effect is found during real-time language comprehension. Some have suggested the effect is found only when the retrieval is triggered (e.g., Gordon et al., 2001; Sekerina et al., 2016; Villata et al., 2018). For example, Sekerina et al. (2016) investigated competition of possible targets during retrieval using a VWP. A sample set of their materials is presented in (4). Four pictures were displayed on the screen: *button* (target), *pen*, *key*, *earring*, and the auditory stimuli in two different conditions were presented. In (4a), multiple objects were plausible targets to be "spotted" (hence the *interference condition*) whereas in (4b), there was only one targeted object (e.g, *button*) that can be "sewed" (i.e., *non-interference condition*). They predicted that targets that share similar features would lead to difficulty in integrating and retrieving the target linguistic items. Specifically, eye gaze towards the target was predicted to be reduced in the presence of another plausible target that shared similar semantic features.

(4) Sekerina et al. (2016)

a. Interference condition (multiple possible targets)
   It was the {button / key / pen / earring} that the maid who returned from vacation **spotted** in the early morning.

b. Non-interference condition (one possible target)
   It was the {button / key / pen / earring} that the maid who returned from vacation **sewed** in the early morning.

They found a significant interference effect, with fewer looks to the target in the interference condition due to the presence of competitors, compared to the non-interference condition. Additionally, this effect was found at the time window when the encoded items needed to be retrieved (at the main verb, e.g., *spotted*, *sewed*). When the verb information

appeared in the audio stream, there were more looks to the target object (e.g., *button*) in the no-competition condition (4b) than in the competition condition (4a).

In contrast, others have found the effect even before the retrieval is initiated (e.g., Acheson & MacDonald, 2011; Kush, Johns, & Van Dyke, 2015; Rich & Wagers, 2020). For example, Rich and Wagers (2020) compared cases where the relevant NPs share semantic features to different extent: (i) high semantic overlap (e.g., "*The knife* that *the sword* was placed near had been recently *sharpened*") and (ii) low semantic overlap (e.g., "*The knife* that *the shirt* was placed near had been recently *sharpened*"). The high-overlap condition led to longer reading times, with the competition effect appearing around *had been recently*, before the relevant semantic information needed to be retrieved and integrated at the verb *sharpened*.

Following prior studies, the current study also identifies two windows. The question is whether the overlap in discourse structure information between two possible targets would lead to an interference effect, specifically, whether the effect will be realized (i) when the second noun phrase is encoded, or (ii) when retrieval is initiated. An overview of the experimental design to test these two hypotheses and the predictions are presented in the following section.

6.3 Overview of experiments and predictions

The key contrast and the time windows of interest are presented in (5). First, in all conditions, both *the singers* and *the violinists* are grammatically plausible antecedents of the pronoun, *their*. Secondly, a distinction is made in the RRC and ARC conditions in the overlap of discourse-structure information between these two antecedents. In the RRC conditions (in (i)), both of the antecedents are part of the main discourse structure; hence, an overlap in discourse-structure information. In the ARC conditions (in (ii)), the second noun phrase (e.g., *the violinists*) is part of the subordinate discourse structure. The two plausible antecedents do not overlap in discourse-structure information. Thirdly, Experiments 7 and

8 had the same experimental design with only minor differences. Experiment 8 included an additional prepositional phrase (e.g., *without reason* before the termination of the relative clause. Experiment 7 had another factor of modifier-length of the head noun.

(5) Experimental items and the two windows of interest in Experiments 7 & 8

a. Experiment 7

(i) Overlap in discourse-structure information between the two antecedents

The singers$_{\text{MAIN}}$ [who admired *the violinists*$_{\text{MAIN}}$]$_{\textbf{window 1}}$ [invited *their* (musical) mentors]$_{\textbf{window 2}}$ to the party. [RRC]

(ii) No overlap in discourse-structure information between the two antecedents

The singers$_{\text{MAIN}}$, [who admired *the violinists*$_{\text{SUBORD.}}$,]$_{\textbf{window 1}}$ [invited *their* (musical) mentors]$_{\textbf{window 2}}$ to the party. [ARC]

b. Experiment 8

(i) Overlap in discourse structure between the two antecedents)

The singers$_{\text{MAIN}}$ [who admired *the violinists*$_{\text{MAIN}}$ without reason]$_{\textbf{window 1}}$ [invited *their* mentors]$_{\textbf{window 2}}$ to the party. [RRC]

(ii) No overlap in discourse structure between the two antecedents)

The singers$_{\text{MAIN}}$, [who admired *the violinists*$_{\text{SUBORD.}}$ without reason,]$_{\textbf{window 1}}$ [invited *their* mentors]$_{\textbf{window 2}}$ to the party. [ARC]

Adopting the approach in Sekerina et al. (2016), eye gaze information, specifically **eye-gaze differences** between the possible targets, will be used to investigate the competition effect. In the case where linguistic elements are encoded with overlapping representations (as in the overlap (RRC) condition), a greater competition effect is expected than when the representations do not overlap (as in the no overlap (ARC) condition). The eye-gaze differences will be smaller when the two antecedents share the same discourse-structure information (i.e., greater competition); the differences are anticipated to be bigger when the

two do not overlap in discourse-structure information. In other words, the prediction on eye-gaze differences is: **ARC (no overlap) > RRC (overlap)**.

Concerning the timeline at which the interference effect is realized, I identify two windows of analysis. Based on prior findings, the windows are defined relative to the retrieval site, i.e., when the linguistic dependency needs to be resolved. The **first window** includes when the second antecedent is encoded. Specifically, it begins from the onset of *who* until the offset of the relative clause. In Experiment 7, it is until the offset of the second NP (*the singers*), and in Experiment 8, it includes the offset of the prepositional phrase (e.g., *without reason* that follows the second antecedent). The **second window** begins from the matrix verb (*invited*) until the offset of the head noun (*mentors*)—the pronoun (*their*) is included in this window.

It is possible that the competition effect, indexed by the eye-gaze differences on the two antecedent images, could arise during the first window, after the second antecedent is encountered (*the violinists*). This would be indicative of an encoding interference effect due to feature overlap between the two encoded NP representations and suggests that the effect could arise prior to retrieval. It is also possible that the effect could manifest only when retrieval is initiated. In this case, the effect should arise during the second window, when the verb or the pronoun is encountered. These two competing possibilities are examined.

To preview the findings, results from two eyetracking studies (VWP) suggest that the overlap in discourse-structure information leads to greater competition between the two plausible targets. Results in Experiment 7 seem to suggest that the effect is realized only when retrieval is initiated. However, in Experiment 8, which was designed to remove the confound of an early interference effect being realized at a later timeline, the effect was found as early as when the second antecedent was encoded. These collectively suggest that the competition effect between linguistic representations due to discourse-structure information overlap can be manifested before retrieval for resolving linguistic dependency.

6.4 Experiment 7

A web-based eyetracking experiment (VWP) is conducted to examine whether the overlap in discourse-structure information between linguistic representations during encoding leads to the interference effect. The key manipulation of the auditory stimuli is repeated below:

(6) a. Overlap in discourse-structure information between the two antecedents
The singers$_{\text{MAIN}}$ [who admired *the violinists*$_{\text{MAIN}}$]$_{\textbf{window 1}}$ [invited *their* (musical) mentors]$_{\textbf{window 2}}$ to the party. [RRC]

b. No overlap in discourse-structure information between the two antecedents
The singers$_{\text{MAIN}}$, [who admired *the violinists*$_{\text{SUBORD.}}$,]$_{\textbf{window 1}}$ [invited *their* (musical) mentors]$_{\textbf{window 2}}$ to the party. [ARC]

We anticipate observing greater eye-gaze differences between the two possible antecedents in the ARC (no-overlap condition) than in the RRC (overlap condition). The empirical question is whether this eye-gaze difference will be realized as early as when the second antecedent is encoded (i.e., in the first window), or only after retrieval is triggered (i.e., in the second window).

6.4.1 Methods

Subjects

Sixty-five native speakers of American English were recruited via an online participant recruitment platform Prolific. The recruitment process for participation was carried out with approval from the IRB from the local institution. Two participants were omitted from the primary analysis because they took part in the study twice, resulting in a total of 63 participants for the analysis (mean age = 30.36; age range: 19–40). The duration of the experiment was around 35 minutes, and the subjects received compensation of 8.00 USD for

their participation.

Material and design

The material was fully crossed in a 2x2 design, varied by Clause (RRC vs. ARC) and Modifier-length (Short vs. Long) (Table 6.1). In the RRC structure, both NPs were part of MAIN information. In the ARC structure, the first NP was part of the MAIN while the second NP was part of the SUBORDINATE discourse structure. The Modifier-length condition was included to offer more time to resolve the pronoun; the Long condition had a modifier (e.g., *musical*) before the head noun. Different vocation names were used for the NPs. All embedded RCs were in a subject-extracted structure. Only *their* was used as the pronoun. The study involved 32 critical and 20 filler items.

Table 6.1: An example of the experimental item for Experiment 7

Clause	Modifier	Auditory stimuli	Discourse status of antecedents {NP 1, NP 2}
RRC	Short	The violinists who admired the singers invited their mentors to the party.	{[MAIN], [MAIN]}
ARC	Short	The violinists, who admired the singers, invited their mentors to the party.	{[MAIN], [SUBORD.]}
RRC	Long	The violinists who admired the singers invited their musical mentors to the party.	{[MAIN], [MAIN]}
ARC	Long	The violinists, who admired the singers, invited their musical mentors to the party.	{[MAIN], [SUBORD.]}

Note. In the example *the violinists* corresponds to NP1 and *the singers* to NP2. Subord. = subordinate. Commas were realized as pauses in auditory stimuli in the visual world paradigm.

Auditory stimuli for both critical and filler items were generated using the open-source Google Text-to-Speech library (`gtts`) in Python. The contrast between ARC and RRC conditions was realized by pauses before and after the RC boundary in the ARC condition, while no pauses were introduced in the RRC condition. Although ARCs exhibit intonation

contour variations, our focus aligns with earlier observations that ARCs typically display prosodic isolation (Dehé, 2014; Dehé & Kavalova, 2007), particularly when compared to RRCs (Watson & Gibson, 2004). The mean duration of pauses for the critical auditory stimuli was 0.327 seconds. The mean length of the auditory stimuli was 7.055 seconds.

Visual stimuli were produced with OpenAI's DALL-E 2 (Ramesh, Dhariwal, Nichol, Chu, & Chen, 2022), an AI system generating images when given a language prompt. For the critical items, each image featured two individuals of the same vocation, created with the prompt: "Drawing of two [VOCATION]s in black and white thick pen line drawing digital art with white background." The vocation names were drawn from the list of NPs utilized in the target items. Similarly, relevant visual stimuli for filler trials were generated using the same method and prompt, with "two [VOCATION]s" replaced by "a [NOUN]," given that filler trials involved only singular nouns.

Procedure

The online visual world paradigm eye-tracking experiment used the PCIbex platform (Zehr & Schwarz, 2018). For the eye-tracking component, PennController version 2.0 was employed in the current study.[3] Participants recruited through Prolific were directed to the online experimental platform, where they were shown an IRB consent page, followed by a language background questionnaire. Then a phase to check the audio volume was presented. On the following page, the subjects were prompted to grant the browser permission to access the web. No video recordings were made during the experiment; only the subjects' eye-gaze information was recorded, and participants were informed of this data collection.

Upon granting access, participants proceeded to the calibration stage. Detailed guidelines were provided regarding the recommended positioning of their face within the camera frame, their posture, and the brightness of their face displayed on the screen. These guidelines were

3. Version 2.1 requires a correct button clicking to move on to the next calibration.

presented with visual examples, adopted from Slim and Hartsuiker (2022). The calibration process involved a button-clicking task. During this initial calibration process, the subjects were asked to look at a green dot (size of 48 px width and 48 px height (approximately 36 pt)) appearing on the screen and to click it when it appeared. A total of 9 dots were presented; the first dot started at the center of the participants' screen (the position of the dot was calculated based on the user's screen size), and the rest appeared in random order, at each edge point of the screen (top left, top center, top right, center left, center right, bottom left, bottom center, and bottom right). The dot appeared on the screen along with a beep sound and stayed on screen for 750 ms, followed by another dot after a delay of 750 ms.[4] The calibration precision was calculated by using the Euclidean distance between the stimulus point, e.g., mid-center dot (X_1, Y_1), and the estimated gaze at the n-th dot (X_2, Y_2). The precision score was calculated as in Equation 6.1.

$$Precision = 100 - \frac{\sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}}{1/2 * WindowHeight} * 100 \tag{6.1}$$

The threshold for acceptable precision score was set as 60, and the participants were asked to repeat the calibration process when they failed to reach the threshold. The study would be aborted if a participant made 3 failed attempts to pass the initial calibration phase. Before the 2nd or 3rd calibration (if they had to), the subjects were given instructions reminding them of the posture and lighting of the camera.

Once the participant passed the calibration, a fixation cross ("+" of 15 vh size) appeared at the screen center for 500 ms.[5] Subsequently, four images in black-and-white line drawings were displayed on each quadrant in a randomized order (Figure 6.1). Each quadrant was a size of 50vw * 50vh (50% of the width and 50% of the height of the participant's screen),

4. In the newer PennController version 2.1, the calibration dots do not stay on the screen for a fixed amount of time. Once participants click on one dot, it disappears and the next dot appears on the screen.

5. "vh" stands for viewport height, and it indicates the height of the browser where a website is displayed. 15 vh corresponds to 15% of the viewport height.

and each image had a size of 20 vh width and 20 vh height. Two of the images depicted the two possible antecedents and the remaining two images were irrelevant distractors.

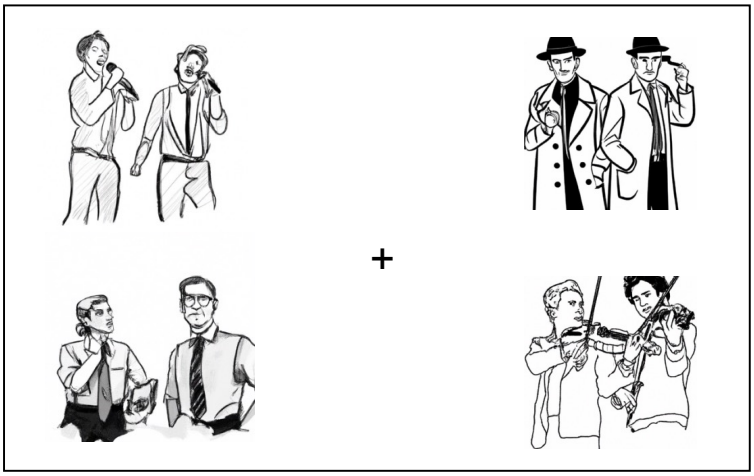


Figure 6.1: An example of the screen in the visual world paradigm experiment.
Note. Clockwise from the top-left: *the singers* (target), *the detectives* (distractor), *the violinists* (target), and *the teachers* (distractor).

After 2200 ms of the image display (image preview time), the auditory stimulus started playing and participants' eye-movement recording also started simultaneously. The eye-movement recording continued for 500 ms after the end of the auditory stimulus. An antecedent selection task followed, with a question prompt asking, "Whose mentors were invited?" Participants were instructed to click on one of the four pictures.

There were between-trial calibration processes as well, a simplified calibration check that was done before each trial. Only one mid-center dot appeared before each trial, and the score was calculated only based on the mid-center dot. The threshold for the between-calibration score was set at 50. Participants below this threshold underwent the initial calibration process (threshold score set at 50 with 2 attempts).

Analysis

Both antecedent selection and eye-movement patterns were analyzed. The performance for the antecedent selection task was used for initial data processing. The trial accuracy was

coded as 1 if one of two plausible NPs were selected and 0 otherwise. Despite having a threshold for removing participants with an accuracy below 85%, all 63 participants passed this criterion. I then examined the eye-movement data. Participants were considered for removal if their eye gaze went outside the quadrants above 50% of the recorded data. No participant was removed based on this criterion. To take into consideration the variations in sampling interval inherent in web-based eyetracking method, I excluded the trials with an average sampling interval that exceeded 100 ms. The mean sampling interval across trials was 33.57 ms (median = 32.82 ms) (Figure 6.2). After applying the sampling interval threshold, 0.01% of the data were removed.

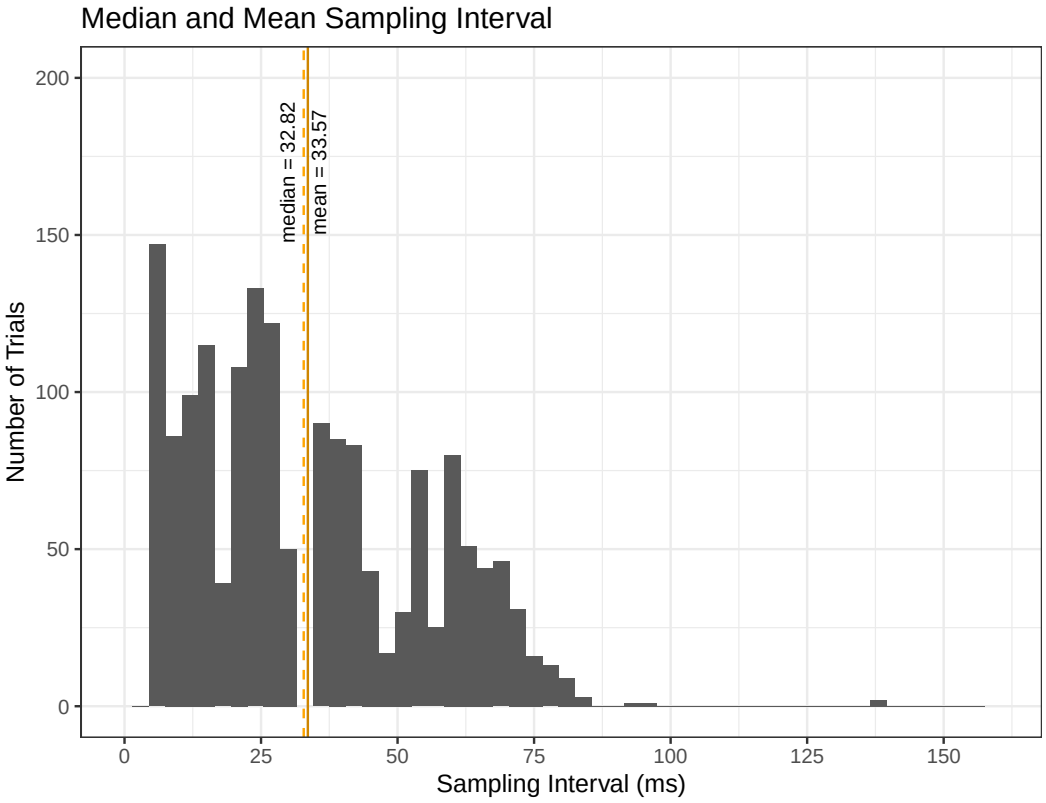


Figure 6.2: Sampling information in Experiment 7.

Subsequently, data points were removed when the eye gaze did not fall within a quadrant (an additional 9.71% data point removal). Additionally, data points with incorrect image selection (i.e., neither of the two NPs selected in the trial) were removed (an additional 1.06%

data point removal). Finally, data points falling outside the designated monitor screen were removed, constituting an additional 0.91% data point removal.

The eye-movement data were aggregated into 100 ms time bins following Slim and Hartsuiker (2022). The proportions of eye-gaze fixations on the four pictures were computed for each time bin (e.g., 0–100 ms, 100–200 ms, ...) across all participants. Taking into account the 200 ms oculomotor delay, the onset of the auditory stimuli was offset by 200 ms.

Given our primary focus on the competition between the two grammatical antecedents, the eye-gaze difference between NP1 and NP2 was computed. This involved subtracting the proportion of eye gazes to NP2 from that of NP1 for each bin. A cluster-based permutation analysis (CPA) (Ito & Knoeferle, 2023; Maris & Oostenveld, 2007) was conducted to examine adjacent time bins that showed significant eye-gaze differences between the two target antecedents ($p < 0.05$). The analysis was employed through the `permutes` package (Voeten, 2023) in R. The eye-gaze difference in the original dataset was examined using mixed-effects linear regression models built with the `lmerTest` package (Bates et al., 2015) in R (R Core Team, 2022). The fixed effects included Clause (RRC vs. ARC), Modifier Length (Short vs. Long), and their interaction, along with a by-subject random effect.[6] Subsequently, the data were randomly permuted 1000 times to establish a null hypothesis distribution. Finally, each cluster within the original dataset was compared against this null hypothesis distribution.

The first window included the onset of *who* until the offset of *the singers*, without including the pauses. The second window included the matrix verb (*invited*) until the offset of the noun (*mentors*) in the Short condition. In the Long condition, the same duration was analyzed in the Long condition—it included the matrix verb, pronoun, the modifier (*musical*), and the beginning of the noun.

6. Regression model for analysis: `eye-gaze_difference ~ clause*modifier + (1|subject)`. Sum coding for the fixed effects: RRC = -0.5 and ARC = 0.5; Short = -0.5 and Long = 0.5.

6.4.2 Results

Antecedent selection preference

In Figure 6.3, the antecedent selection preference ratio is depicted. Irrespective of conditions, NP1 exhibited a higher preference, approximately 65%, while NP2 showed a preference of around 35%. Two of the distractors were rarely selected.

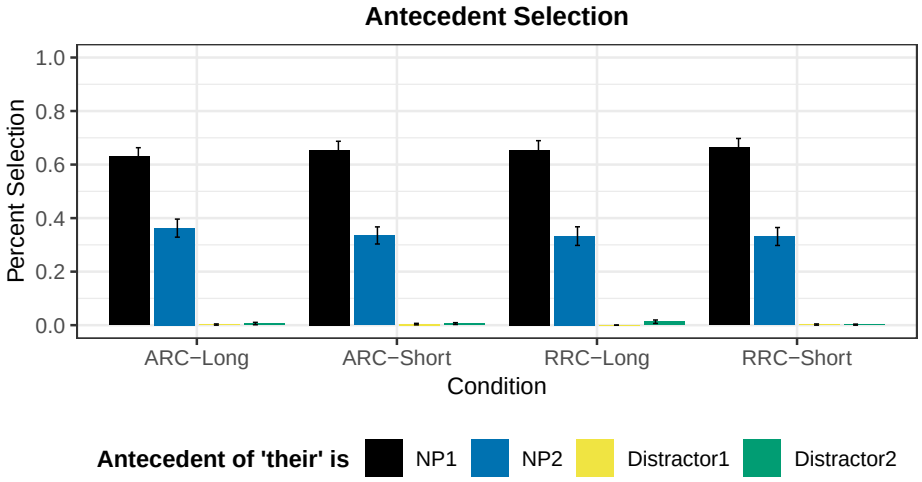


Figure 6.3: Antecedent selection preference in Experiment 7.
Note. Error bars indicate standard errors of the mean.

Eye-gaze ratio

Figure 6.4 shows the proportion of eye-gaze onto figures at each time bin, averaged across all participants.

The first window The mean eye-gaze ratio in the first window is shown in Figure 6.5. The difference between eye-gaze on NP1 and NP2 is illustrated in Figure 6.6. The mean eye-gaze difference between NP1 and NP2 is shown in Figure 6.6. The CPA suggested that no statistically significant clusters were found ($ps > 0.05$).

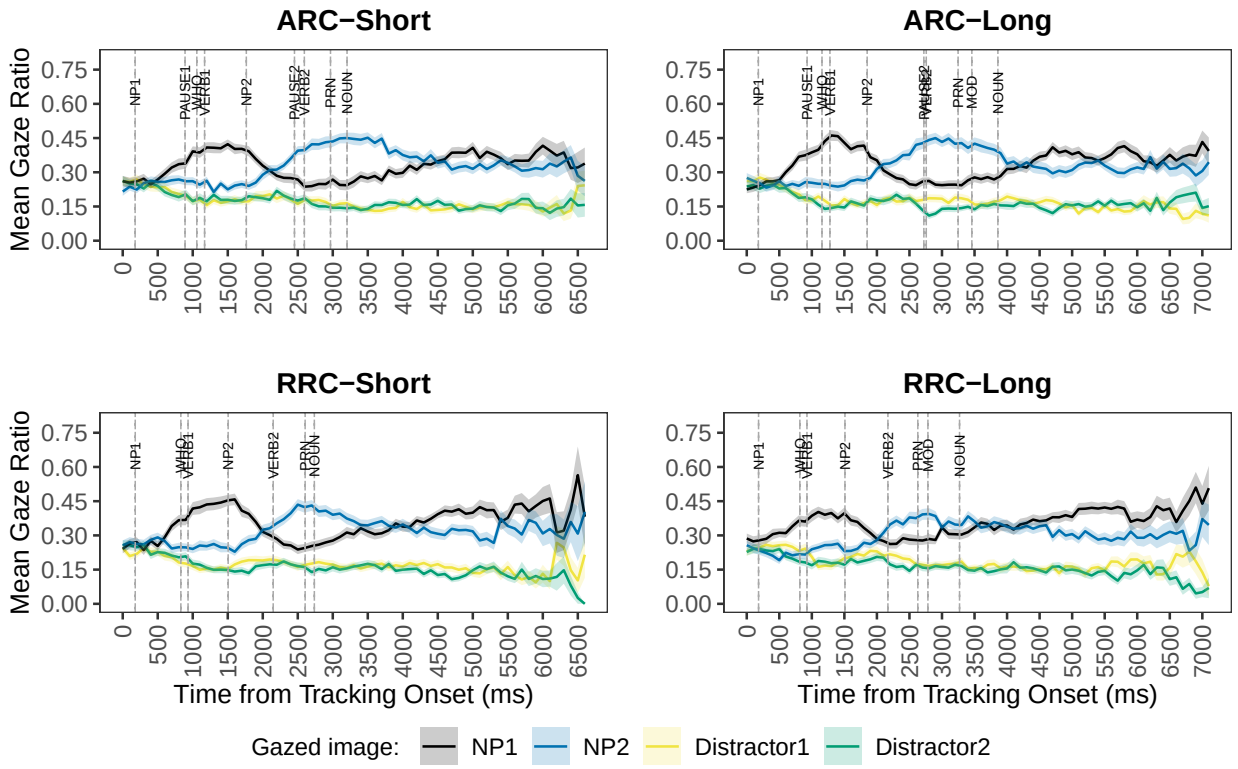


Figure 6.4: Eye-gaze ratio in Experiment 7.

Note. The labels above the vertical lines denote the average onset of each word across items. NP1 corresponds to *the violinists*; VERB1 to *admired*; NP2 to *the singers*; VERB2 to *invited*; PRN (pronoun) to *their*; MOD (modifier) to *musical*; NOUN to *mentors*; PAUSE1/2 represents pauses inserted in the ARC condition, as in the trial "The violinists, who admired the singers, invited their (musical) mentors to the party." Shaded ribbons surrounding the lines indicate the standard errors of the mean.

The second window The mean eye-gaze ratio in the second window is shown in Figure 6.7. The difference between eye-gaze on NP1 and NP2 is illustrated in Figure 6.8. A significant cluster (0 ms–800 ms) for the Clause effect ($p < 0.001$) was found. This suggests that the difference (ARC > RRC) started from the verb onset (Figure 6.8). No significant clusters were found for other effects.

6.4.3 Discussion

In the pronoun-antecedent selection task, there was a preference for the first NP as the antecedent of *their*, regardless of all the conditions. This replicates the observation made with

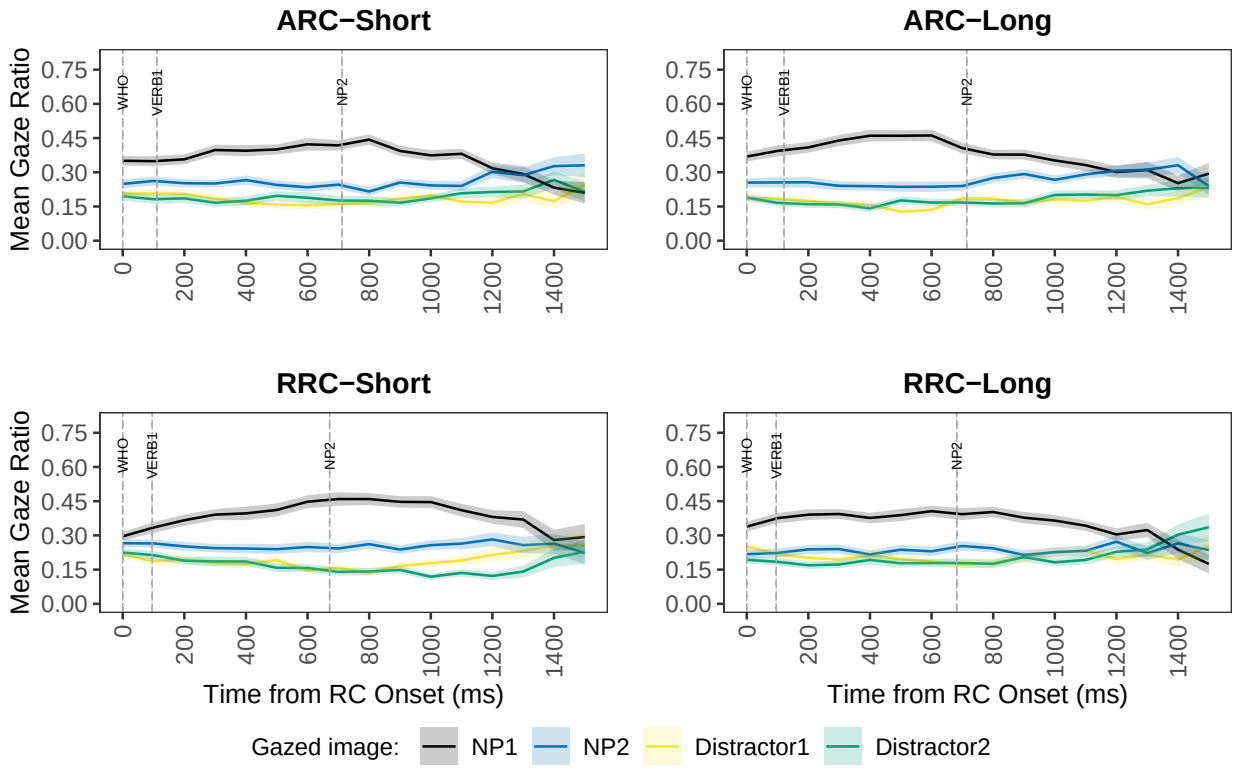


Figure 6.5: Eye-gaze ratio in the RC window in Experiment 7.

Note. The labels on top of the vertical lines indicate the average onset of the word across items. WHO = *who*; VERB1 = *admired*; NP2 = *the singers*, as in the trial "The violinists(,) who admired the singers(,) invited their mentors to the party." Shaded ribbons surrounding the lines indicate the standard errors of the mean. The X-axis is time-locked to the onset of "WHO" across items. Shaded ribbons around lines indicate the standard errors of the mean.

the sentence-medial RC condition in Experiment PRN-SPP-01, even the ratio of preference, where NP1 was preferred around 60% and NP2 around 30%. The preference for the NP1 in the selection task is also hinted at in the eye-gaze results. For instance, in all four subfigures in Figure 6.4, the eye-gaze towards the NP1 rises around 1000 ms after the onset of the final word of the sentence, the head noun. This is an overall pattern across all four conditions, which again matches the results in the antecedent selection task results. This global preference for the NP1 underscores a general inclination to choose the subject of the same clause when faced with ambiguous antecedents, as noted in J. E. Arnold (2010). The consistent preference for NP1 across conditions is noteworthy, suggesting that the observed patterns in the eye-gaze results cannot be solely attributed to subjects' global preferences

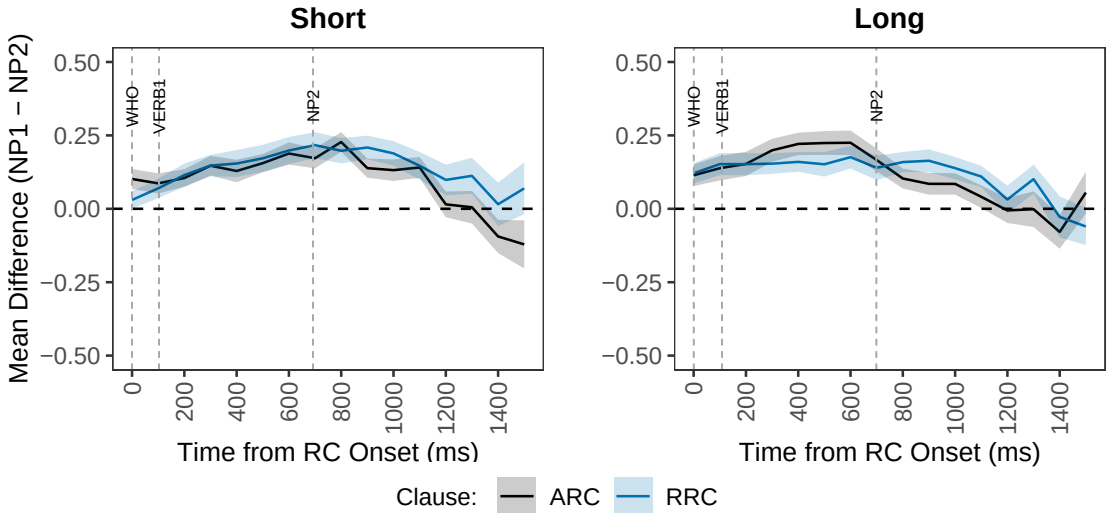


Figure 6.6: Eye-gaze ratio difference between NP1 and NP2 in the RC window in Experiment 7.

Note. The labels on top of the vertical lines indicate the average onset of the word across items. WHO = *who*; VERB1 = *admired*; NP2 = *the singers*, as in the trial "The violinists(,) who admired the singers(,) invited their mentors to the party." Shaded ribbons surrounding the lines indicate the standard errors of the mean. The X-axis is time-locked to the onset of "WHO" across items. Shaded ribbons around lines indicate the standard errors of the mean.

for a specific antecedent.

The analysis of eye-gaze results revealed a competition effect due to overlap in discourse structure information, leading to smaller eye-gaze ratio differences between possible antecedents (difference: RRC < ARC). The effect began only after the onset of the matrix verb.

It is worthwhile to note that the competition effect started from the matrix verb (e.g., *invited*). There are two possible accounts for why the effect would have been realized as early as the matrix verb window, which follows below. One possibility is due to the reactivation of previously encoded linguistic representation at the retrieval site. If [+subject] and [+animate] are key retrieval cues for resolving the subject-verb dependency, linguistic representations with these features can be activated at the retrieval site. In the current experiment, the matrix subject (e.g., *the violinists*), associated with both the [+subject] and [+animate] cues, and the embedded object (e.g., *the singers*), associated with the [+animate]

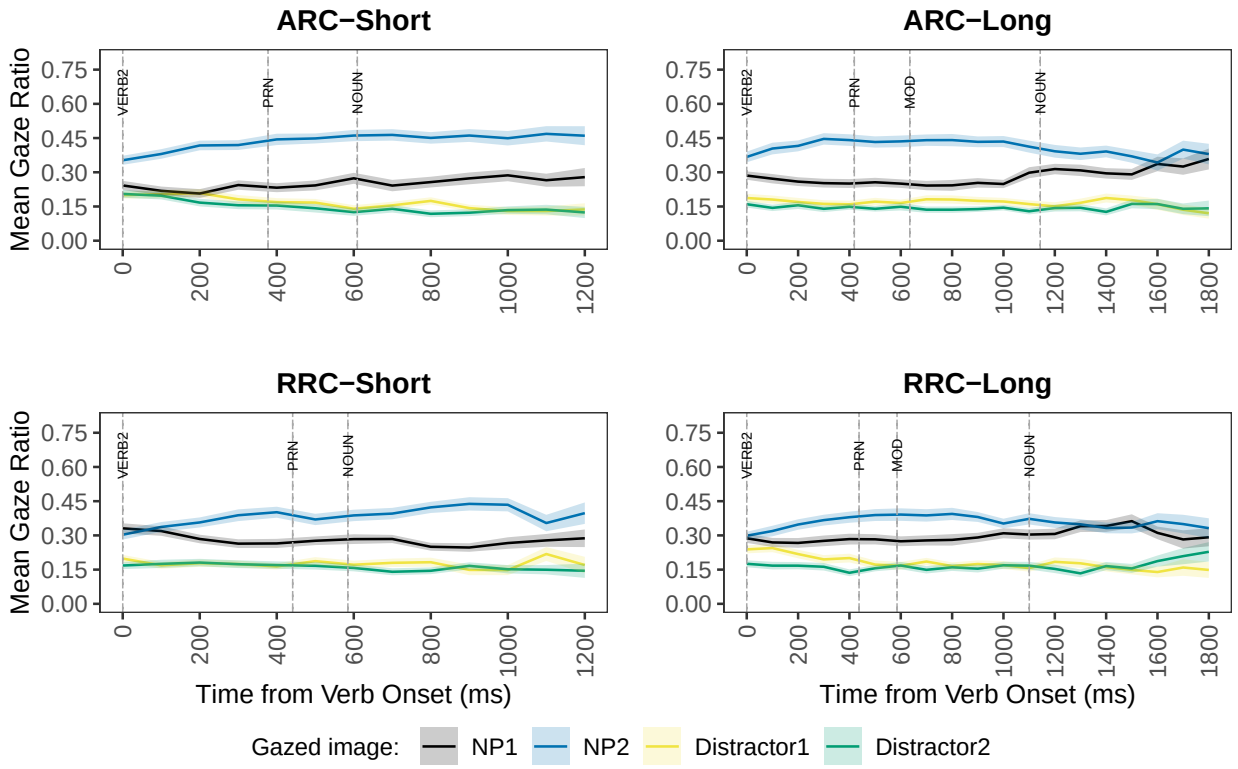


Figure 6.7: Eye-gaze ratio in the Pronoun window in Experiment 7.
Note. The labels on top of the vertical lines indicate the average onset of the word across items. VERB2 = *invited*; PRN (pronoun) = *their*; MOD (modifier) = *musical*; NOUN = *mentors*, as in the trial "The violinists(,) who admired the singers(,) invited their (musical) mentors to the party." Shaded ribbons surrounding the lines indicate the standard errors of the mean. The X-axis is time-locked to the onset of VERB2 across items. Shaded ribbons around lines indicate the standard errors of the mean.

cues, can be activated. Even though discourse structure information is not necessarily a retrieval cue at the matrix verb, the re-activation of the previously encoded linguistic representations associated with the retrieval cues could have led to the distinction between the two NPs with distinct discourse structure information.

Alternatively, the presence of the competition effect at the matrix verb region could be a reflection of a delayed realization of the effect in the previous window. Returning to Sekerina et al.'s work, the authors found an interference effect due to overlap in semantic features, evidenced by smaller eye-gaze differences between retrieval targets. They found the effect in the retrieval beginning from the matrix verb window (e.g., *spotted*, *sewed*) (an example

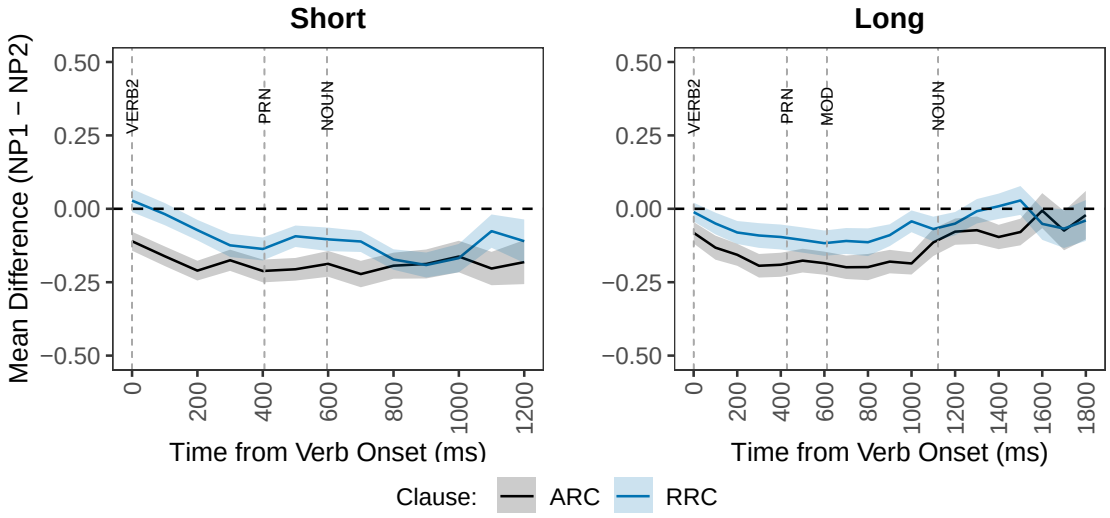


Figure 6.8: Eye-gaze ratio difference between NP1 and NP2 in the Pronoun window in Experiment 7.

Note. The labels on top of the vertical lines indicate the average onset of the word across items. VERB2 = *invited*; PRN (pronoun) = *their*; MOD (modifier) = *musical*; NOUN = *mentors*, as in the trial "The violinists(,) who admired the singers(,) invited their (musical) mentors to the party." Shaded ribbons surrounding the lines indicate the standard errors of the mean. The X-axis is time-locked to the onset of VERB2 across items. Shaded ribbons around lines indicate the standard errors of the mean.

material repeated in (7). However, it is worthwhile to note that the effect they observed spanned until the following prepositional phrase (e.g., *in the early morning*). The authors explain that this could be a reflection of an end-of-the-sentence, wrap-up effect, but crucially, a reflection of the participants' search for the target(s).

(7) A set of sample material in Sekerina et al. (2016)

a. Interference condition (multiple possible targets)

It was the {button / key / pen / earring} that the maid who returned from vacation **spotted** in the early morning.

b. Non-interference condition (one possible target (*button*))

It was the {button / key / pen / earring} that the maid who returned from vacation **sewed** in the early morning.

As their experimental design and results suggest, it is possible that the effect of interest spans to the following window. However, the current experimental design makes it difficult to tease apart such a delayed effect given the position of the second NP. The offset of the second NP aligns with the rightmost boundary of the RC, which marks the termination of subordinate discourse in the case of the ARC condition. As the transition from one discourse to another occurs immediately after the second NP, the observed effect in the current experiment could reflect an effect that was present from the early, first window around the second NP.

To address this confound, in the following experiment, a buffer window is included before the RC boundary. Specifically, a prepositional phrase (PP) is added after the second NP before the matrix verb. This manipulation is made to offer extra time to encode the distinct discourse structure information. If the competition effect is realized as early as during the encoding stage, we expect to see the eye-gaze differences before the second window.

6.5 Experiment 8

This experiment aims to tease apart between the two possibilities of understanding the competition effect realized at the retrieval window. There are existing findings that longer encoding time can make representations stronger and more salient in the memory since it leads to more attentional resources and a higher activation level of the representation (e.g., Karimi, Diaz, & Wittenberg, 2020). Hence, it is possible that the transition to a different discourse right after the second NP in the RC was not enough time to encode the second NP. To address this possibility, an additional linguistic buffer, i.e., a prepositional phrase, is added before the termination of the RC, before moving to the matrix verb window. If the competition effect is realized as early as during the encoding stage given the additional buffer, we expect to observe the eye-gaze differences before the second window (i.e., before the matrix verb).

6.5.1 Methods

Subjects

A total of 149 native speakers of American English were recruited through the online participant recruitment platform Prolific. The recruitment process adhered to the approval granted by the IRB at the local institution. Some participants were unable to complete the experiment, and I excluded these individuals from the main data analysis, resulting in 108 participants (mean age = 30.37; age range: 18-–40). The experiment duration was approximately 35 minutes, and participants received compensation of 8.50 USD for their participation, including those who could not complete the experiment.

Material and design

Experiment 8 utilized the same set of 32 items and 20 filler items as Experiment 7, with the addition of a PP before the end of the RC. Table 6.2 illustrates a comparison of the two experimental designs. Since there was no observable length effect of the noun-modifier in the previous experiment (Experiment 7), the Length factor is not included in the current experiment.

Procedure

An online visual world paradigm eye-tracking experiment was conducted, implemented with PCIbex (version 2.0) (Zehr & Schwarz, 2018). Participants were recruited through Prolific. The experimental method and platform were identical to those used in Experiment 7.

As for calibration, just as in Experiment 7, I set the calibration score threshold as 60 for the initial stage, and the participants had three opportunities to calibrate. Participants who were not able to pass the initial calibration stage were redirected to a different experiment. However, I set the between-trial calibration score as 30, a lower score than the threshold in

Table 6.2: Comparison of the experimental items in Experiments 7–8

Experiment	Clause	Sentence	Discourse status of antecedents {NP 1, NP 2}
Exp 7	RRC	The violinists who admired the singers invited their (musical) mentors to the party.	{[MAIN], [MAIN]}
Exp 7	ARC	The violinists, who admired the singers, invited their (musical) mentors to the party.	{[MAIN], [SUBORD.]}
Exp 8	RRC	The violinists who admired the singers without reason invited their mentors to the party.	{[MAIN], [MAIN]}
Exp 8	ARC	The violinists, who admired the singers without reason, invited their mentors to the party.	{[MAIN], [SUBORD.]}

Note. In the example *the violinists* corresponds to NP1 and *the singers* to NP2. Subord. = subordinate.

Experiment 7, which had a threshold of 50. If the between-trial calibration score was below 30, the participants were instructed to go through the calibration process again; if they failed to pass the score of 30 3 times in a row, they were given a completion code for compensation and were instructed to exit from the experiment.

The visual presentation of the pictures and fixation cross were presented in the same way as in Experiment 7. There was 500 ms of '+' fixation time, and after the fixation cross disappeared, the four images were presented for 2200 ms. Then the auditory stimuli were played, and there was an additional 500 ms overspill after the offset of the audio. After 200 ms, the antecedent selection question appeared. The mouse cursor was activated right after the question. Unlike Experiment 7, where the mouse cursor showed only when it hovered above the pictures, the cursor was visible on the entire screen so that the participants were able to locate the cursor. And similar to Experiment 7, there was no time constraint on the picture selection task.

Analysis

The task performance was examined based on their task response. Their task accuracy was measured by the mean task performance score, where the correct trial was coded as 1 (correct) if one of the two target pictures was selected and 0 (incorrect) if one of the two distractor pictures was chosen. I set a threshold of 85% so that participants with an accuracy score below the threshold should be removed—no participants were removed. I then examined the eye-gaze track loss, with the criterion of removing participants whose eye-gaze fell outside of the quadrants more than half of the recording time points—all participants passed the threshold. Then the average calibration score throughout the trials was calculated. Participants whose mean calibration score was 50 or below were excluded; 8 were removed after this process.

Data points were additionally removed based on the following criterion: a criterion of removing data points with a sampling interval over 100 ms. The mean sampling interval was 46.45 ms (median = 42.82 ms) (Figure 6.9). A total of 0.04% of the data points were removed due to the high sampling interval. Additionally, data points were removed that had no image gaze (additional 8.19%), that selected the distractor (incorrect) images (additional 1.11%), that had eye-gazes off the monitor screen (additional 0.97%), and that had calibration scores below 50 (13.9%). Applying these criteria, a total of 22.62% of the initial data points were removed.

Similar to the analysis in Experiment 7, I followed the analysis method in Slim and Hartsuiker (2022) where the data points were binned for every 100 ms interval to accommodate the low sampling rate of the data collected online. To adjust the oculomotor delay, the data points were analyzed 200 ms after the onset of the auditory stimuli, following the conventions in the visual world paradigm experiments.

Two windows of analysis were identified (see ((8)). The first window started from the onset of the relative pronoun (*who*), until the offset of the PP (e.g., *without reason*). The

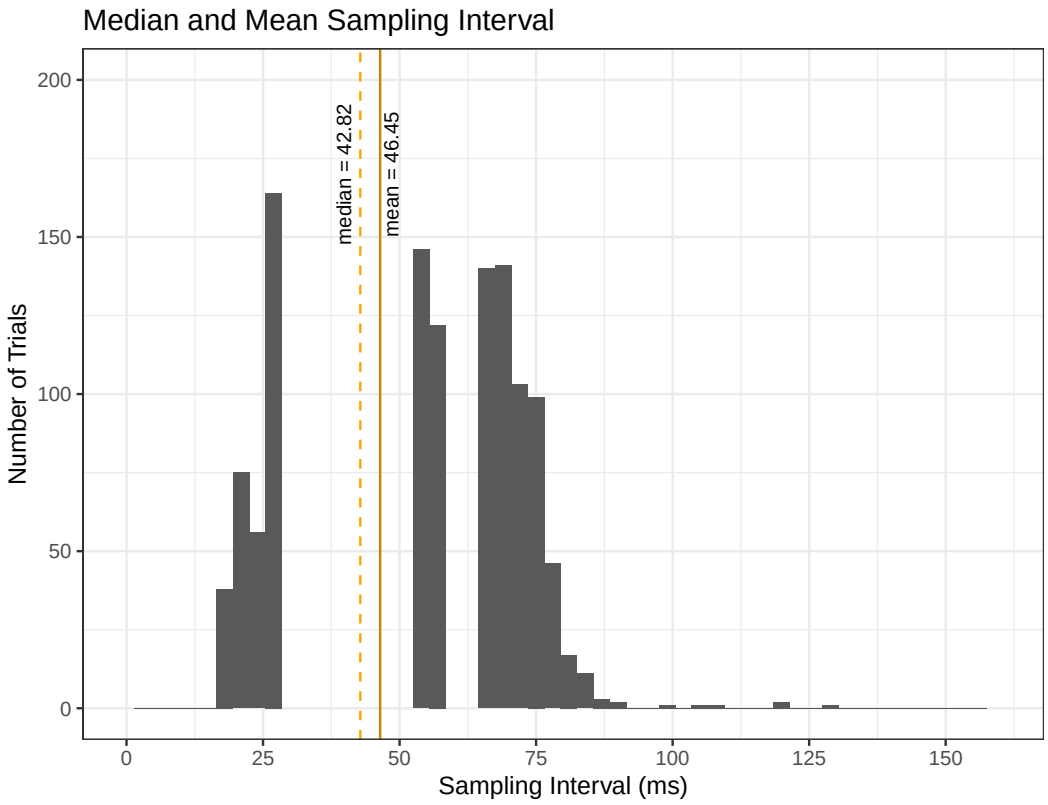


Figure 6.9: Sampling information in Experiment 8.

second window included the onset of the matrix verb (e.g., *invited*), to the offset of the head noun (e.g., *mentors*).

(8) a. Overlap in discourse structure between the two antecedents)
The singers$_{\text{MAIN}}$ [who admired *the violinists*$_{\text{MAIN}}$ without reason]$_{\textbf{window 1}}$ [invited *their* mentors]$_{\textbf{window 2}}$ to the party. [RRC]

b. No overlap in discourse structure between the two antecedents)
The singers$_{\text{MAIN}}$, [who admired *the violinists*$_{\text{SUBORD.}}$ without reason,]$_{\textbf{window 1}}$ [invited *their* mentors]$_{\textbf{window 2}}$ to the party. [ARC]

The same approach of cluster-based permutation analysis (CPA) (Ito & Knoeferle, 2023; Maris & Oostenveld, 2007) was applied as in Experiment 7. The difference between the

eye-gaze ratio between NP1 and NP2 was calculated and used for the main analysis. A null hypothesis distribution was established through 1000 times of random permutation with the observed data. The fixed effect in this null distribution was compared with the clusters within the original dataset to find the significant clusters. These processes were done through the `permutes` package (Voeten, 2023) implemented in R (R Core Team, 2022). Mixed-effects linear regression models were built to analyze the original dataset, and this was done using the `lmerTest` packaged in R (R Core Team, 2022). The difference value (= eye-gaze on NP1 − eye-gaze on NP2) was included as the dependent variable. Clause factor was included as a fixed effect in the model, and a by-subject random intercept was also included in the model.[7]

6.5.2 Results

Antecedent selection preference

The antecedent selection task results are illustrated in Figure 6.10. The preference for NP1 was around 70%, and 30% for NP2, irrespective of the clause type. The two other distractors were rarely chosen as plausible antecedents.

Eye-gaze ratio

The mean proportion of eye-gaze, averaged across all participants, at each time bin, is presented in Figure 6.11.

The first window Figure 6.12 shows the mean eye-gaze ratio during the first window. The difference between eye-gaze on NP1 and NP2 is illustrated in Figure 6.13. A significant cluster was found in 1700 ms–1900 ms and 2100 ms–2300 ms after the onset of the RC

7. Regression model for analysis: `eye-gaze_difference ~ clause + (1|subject)`. Sum-coding of fixed effects: RRC = -0.5; ARC = 0.5.

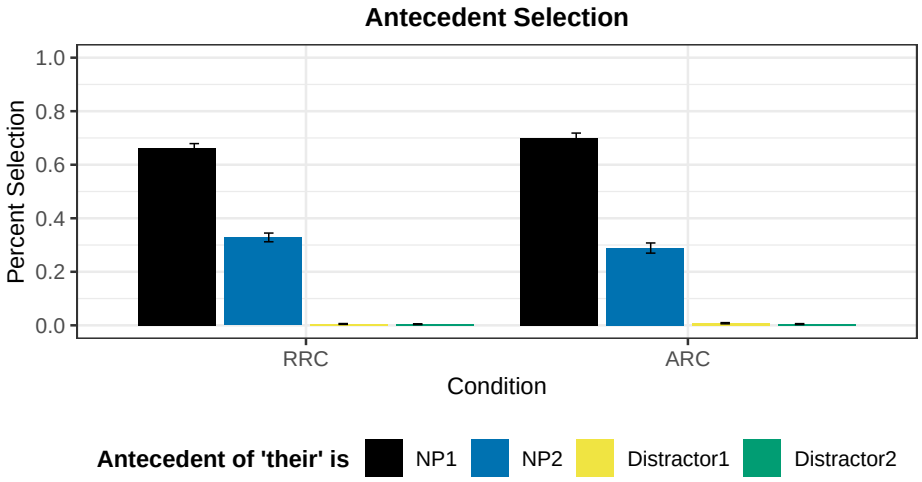


Figure 6.10: Antecedent selection preference in Experiment 8.
Note. Error bars indicate standard errors of the mean.

($p < 0.01$), which suggests that the effect (i.e., ARC > RRC in eye-gaze difference) started around the onset of the PP.

The second window Figure 6.14 shows the mean eye-gaze ratio on each of the four pictures presented on the screen. Crucially, Figure 6.15 presents the difference between the eye gaze on NP1 and the eye gaze on NP2. No significant cluster groups were identified ($ps > 0.1$). The ARC vs. RRC contrast did not lead to eye-gaze differences during retrieval.

6.5.3 Discussion

Beginning with the antecedent selection task, NP1 was most preferred as the antecedent of the pronoun, replicating the findings in previous experiments. The selection proportion between NP1 and NP2 was similar to the earlier experiments, where NP1 was chosen 65% of the time and NP2 around 35%. This selection preference was uniform between clause types, again showing a preference for a subject and the subject within the same clause (e.g., J. E. Arnold, 2010). This again suggests that the observed eye-gaze differences cannot be explained solely by antecedent preference.

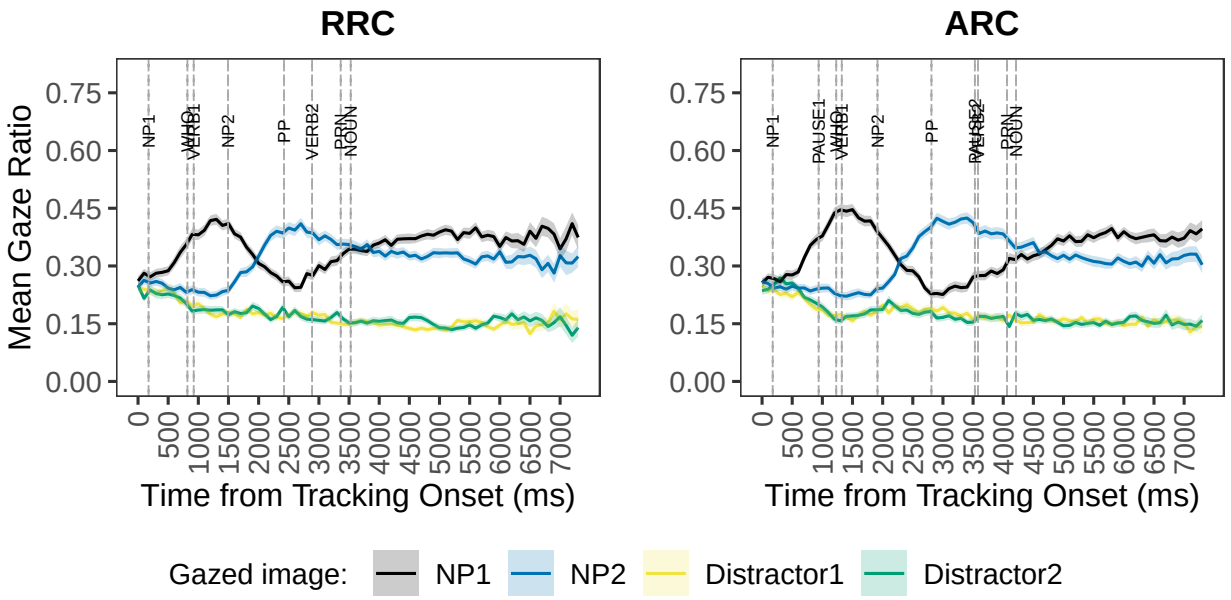


Figure 6.11: Eye-gaze ratio in Experiment 8.

Note. The labels above the vertical lines denote the average onset of each word across items. NP1 corresponds to *the violinists*; VERB1 to *admired*; NP2 to *the singers*; PP to *without reason*; VERB2 to *invited*; PRN (pronoun) to *their*; NOUN to *mentors*; PAUSE1/2 represents pauses inserted in the ARC condition, as in the trial "The violinists, who admired the singers without reason, invited their mentors to the party." Shaded ribbons surrounding the lines indicate the standard errors of the mean.

Regarding the eye-gaze difference, the RRC condition ({MAIN, MAIN} overlap) had smaller eye-gaze differences than the ARC condition ({MAIN, SUBORDINATE}). The purpose of the current experiment was to understand whether the eye-gaze difference beginning at the second window in Experiment 7 was a reflection of a delayed competition effect. Here we observed the effect at the first window during the PP, which is after the second NP but before the retrieval site. This suggests that the overlap in discourse structure information leads to a competition effect after the overlapping representation is encoded. Additionally, it indicates that the effect in the second window in Experiment 7 was a delayed realization of an encoding interference effect.

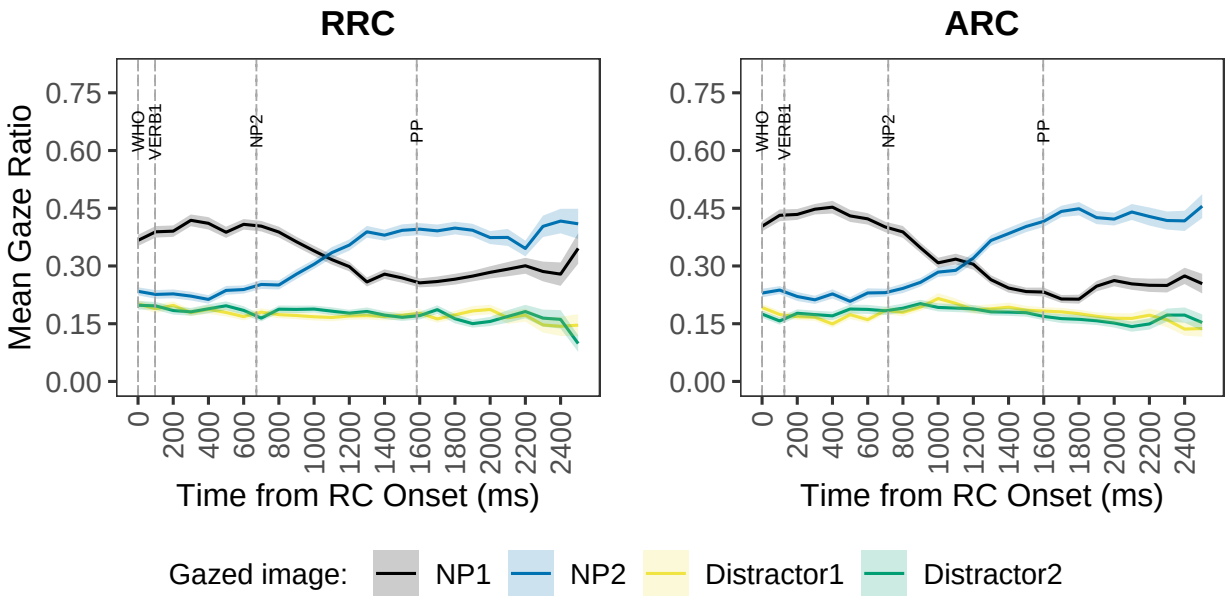


Figure 6.12: Eye-gaze ratio in the first window in Experiment 8.
Note. The labels on top of the vertical lines indicate the average onset of the word across items. WHO = *who*; VERB1 = *admired*; NP2 = *the singers*; PP = *without reason*, as in the trial "The violinists(,) who admired the singers(,) invited their mentors to the party." The X-axis is time-locked to the onset of "WHO" across items. Shaded ribbons around lines indicate the standard errors of the mean.

6.6 General discussion

6.6.1 Summary of the findings

The two sets of eyetracking experiments were conducted to explore (i) whether the overlap of discourse structure information leads to an encoding interference effect, and (ii) how early/late that effect is realized. An intra-sentential anaphoric dependency was examined, wherein a pronoun (*their*) had two possible antecedents. In one condition, the two antecedents were both part of the main discourse structure. Hence, the two antecedents overlapped in discourse-structure information, i.e., both [MAIN]. In the other condition, the two antecedents did not share the same discourse-structure information: one was part of [MAIN] while the other was part of [SUBORDINATE] discourse structure. The overlap state of the two antecedents was manipulated by the contrast in the RRC (overlap) vs. ARC (no over-

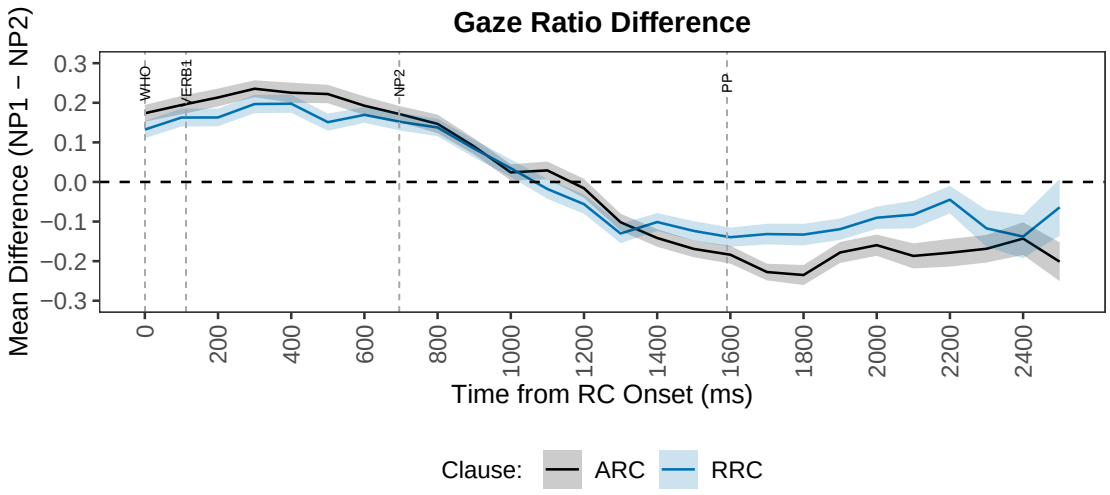


Figure 6.13: Eye-gaze ratio difference between NP1 and NP2 in the first window in Experiment 8.

Note. The labels on top of the vertical lines indicate the average onset of the word across items. WHO = *who*; VERB1 = *admired*; NP2 = *the singers*; PP = *without reason*, as in the trial "The violinists(,) who admired the singers(,) invited their mentors to the party." The X-axis is time-locked to the onset of "WHO" across items. Shaded ribbons around lines indicate the standard errors of the mean.

lap) structure. It was predicted that competition for the shared information between the linguistic entities would lead to a degraded representation. In a visual world paradigm, this degraded representation would be reflected as smaller eye-gaze differences between the two antecedents.

Next, I also addressed the question of how early the competition effect would be realized. There have been mixed results in the literature as to whether an encoding interference effect is realized only when retrieval is triggered. To address this issue, two windows of interest for analysis have been identified. The first window included the window that involved encoding the second antecedent. The second window included the retrieval site. It was an empirical question whether the effect would be realized before the retrieval site, that is, as early as when the second antecedent, which has overlapping linguistic representation as the first antecedent, is encoded.

The windows with significant clusters in each experiment are summarized in Table 6.3.

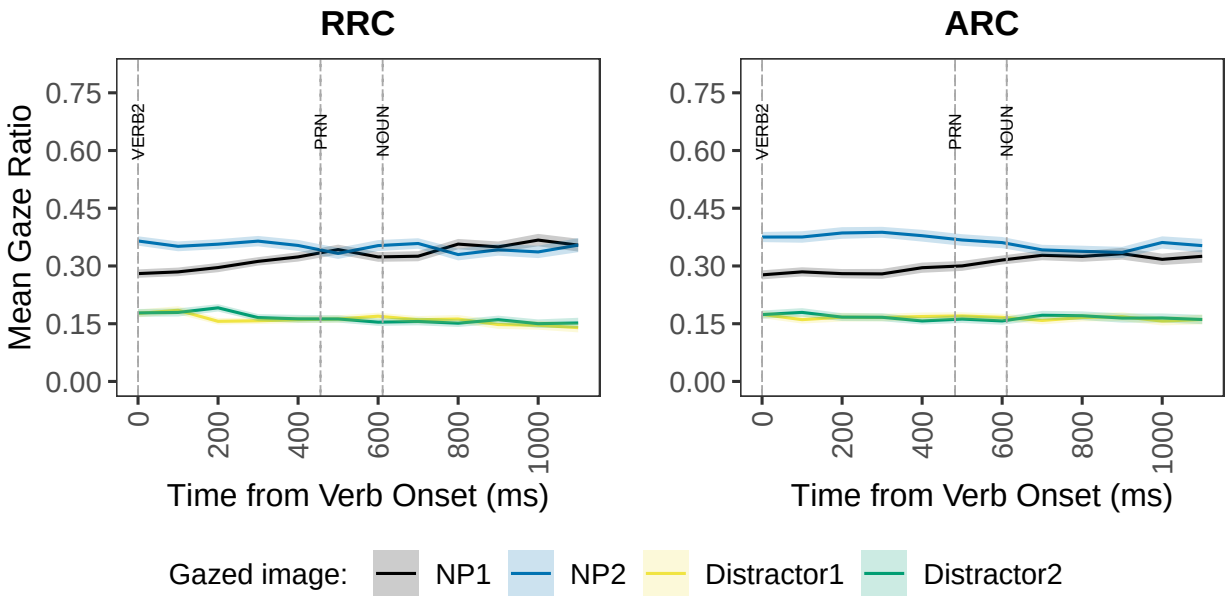


Figure 6.14: Eye-gaze ratio in the second window in Experiment 8.
Note. The labels on top of the vertical lines indicate the average onset of the word across items. VERB2 = *invited*; PRN (pronoun) = *their*; NOUN = *mentors*, as in the trial "The violinists(,) who admired the singers(,) invited their mentors to the party." The X-axis is time-locked to the onset of VERB2 across items. Shaded ribbons around lines indicate the standard errors of the mean.

The findings collectively suggest that when the encoded representations share the same discourse structure information (as in the RRC condition), it results in a competition effect, compared to when they have distinct discourse information (as in the ARC condition). This was shown by the more pronounced eye-gaze differences between the two encoded NPs in the ARC condition compared to the RRC condition. Next, regarding the time window issue, In Experiment 7, the eye-gaze difference was found only in the second window, right after the onset of the first word, the matrix verb; this effect lasted around 800 ms. This seemed to suggest that the overlap effect is realized when retrieval of prior linguistic representations is triggered. However, when more time was given to encode the second NP before the end of the RC in Experiment 8, the eye-gaze difference started as early as the first window of interest, specifically after the second NP and before the termination of the RC. The effect lasted around 600 ms (with a gap of around 200 ms of time window in between). This suggests

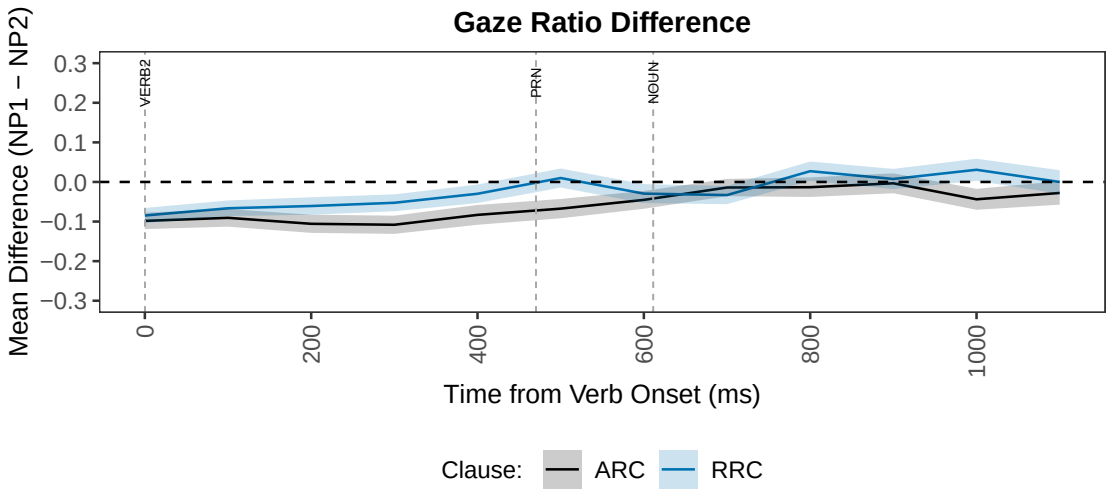


Figure 6.15: Eye-gaze ratio difference between NP1 and NP2 in the second window in Experiment 8.

Note. The labels on top of the vertical lines indicate the average onset of the word across items. VERB2 = *invited*; PRN (pronoun) = *their*; NOUN = *mentors*, as in the trial "The violinists(,) who admired the singers(,) invited their mentors to the party." The X-axis is time-locked to the onset of VERB2 across items. Shaded ribbons around lines indicate the standard errors of the mean.

that the encoding interference effect can be realized even before retrieval is triggered. The effect in Experiment 7 could be interpreted as the same effect we observed in Experiment 8 but simply overlapped with the onset of the retrieval site.

As recent studies find and previously discussed, there can be delayed effects in the web-based eye-tracking experiment (Degen et al., 2021; Lee & Kaiser, 2022; Semmelmann & Weigelt, 2018; Slim & Hartsuiker, 2022; Vos et al., 2022). One may question whether the interpretation of the findings would be affected by such characteristics of the current method. Yet, this is unlikely. The competition effect was already realized around the onset of the PP in Experiment 8. Even if there was a delayed effect incurred by the low spatiotemporal resolution of the web-based eyetracking method, the effect would still show up before the retrieval is triggered, if not as early as around the second NP. Hence, the findings together suggest that the competition effect due to overlap in linguistic information can be manifested as early as during encoding, in line with earlier work that reported an encoding interference

Table 6.3: Comparison of eye-gaze differences in Experiments 7–8 in each window

	Window 1	**Window 2**
Exp. 7	who admired the singers	invited their (musical) mentors
	n.s.	RRC < ARC (800 ms from verb onset)
	Window 1	**Window 2**
Exp. 8	who admired the singers without reason	invited their mentors
	RRC < ARC (for about 600 ms from PP onset)	*n.s.*

Note. The effect of discourse-structure information overlap (RRC (overlap) vs. ARC (no overlap)) on eye-gaze differences on two possible antecedents is compared. A cluster-based permutation analysis is conducted for each window. *n.s.* = no significant clusters were found.

effect near but before retrieval (e.g., Acheson & MacDonald, 2011; Kush, Johns, & Van Dyke, 2015; Rich & Wagers, 2020).

6.6.2 Visual world paradigm for pronoun resolution

A large body of work has explored the real-time processing of pronoun resolution, particularly focusing on the time course of resolving pronouns. Different methods have been used to understand anaphoric dependency resolution, ranging from priming paradigm (e.g., Nicol & Swinney, 1989), picture selection tasks (e.g., Kaiser et al., 2009), and more time-sensitive measures such as self-paced reading tasks (e.g., Badecker & Straub, 2002), eyetracking-while-reading tasks (e.g., Sturt, 2003), visual world paradigm with eyetracking method (e.g., J. E. Arnold et al., 2000) and ERP measures (e.g., Nieuwland & Van Berkum, 2006). Some have used multiple methods in the same study to complement the strengths of each task.

A potential concern can arise as to whether a visual world paradigm is suitable for investigating pronoun resolution. Unlike cases where objects are presented on a screen and participants make eye gaze shifts to entities that can be integrated with the verb (e.g., an auditory stimulus in Tanenhaus et al. (1995): "Put the apple on the towel in the box"), the integration of an antecedent for a pronoun can be a *passive* task that may not be necessarily

reflected on eye gazes.

It is worthwhile to note though that the visual world paradigm experiments in the current study accompanied an *active* task of pronoun resolution; hence, not necessarily *passive*. Although the experiments did not involve an explicit goal-targeted task such as moving the apple on the screen to a different place in the visual world using a mouse cursor, for example, participants were asked to choose the most plausible referent of the antecedent. This antecedent identification question followed after each trial even in filler trials, and this type of task over trials would encourage participants to be engaged in encoding the referents and retrieving them for pronoun resolution. Kaiser et al. (2009), for example, similarly in a pronoun resolution experiment using a visual world paradigm, used the same type of picture-selection task, and were able to find evidence for pronoun resolution. In fact, some studies show that participants engage in pronoun resolution even when the task does not involve picture selection. For example, in Han et al. (2021), rather than a picture-selection task, a comprehension question about the auditory stimuli was given after each trial. Even in this task setup, the authors found evidence of real-time pronoun resolution using a visual world paradigm.

Furthermore, it is less likely that participants did not gaze at the referent even when they were processing it to resolve the pronoun. In both eyetracking experiments, there was a preference for selecting the matrix subject (NP1) as the referent of the pronoun. This preference was clearly reflected in the eye-gaze data (see Figures 6.4 and 6.11), where regardless of the condition (be it RRC or ARC condition), the eye gaze towards the matrix subject (NP1) increased after the onset of the pronoun (around 500–1000 ms post-onset). This would suggest that the current design captures the real-time pronoun resolution process.

On the timing of pronoun resolution

It is worthwhile to note that the eye-gaze differences between ARC vs. RRC were not observed in the second window of analysis. This may be unexpected when the self-paced reading task experiments (in Chapter 5) are considered. In those experiments, the RRC vs. ARC contrast led to different reading times at the retrieval site, specifically when the pronoun had to be resolved. Why is the effect realized around the pronoun resolution site in the self-paced reading task experiments but not in the eyetracking experiment?

There are two possible explanations. One is to attribute to the nature of the visual world paradigm that, unlike reading experiments, the parser is not engaged in resolving the pronoun. However, as I discussed in the previous section, this is unlikely. Even if it was not specifically at the second window of analysis, the eye gaze shifted towards the first NP at the end of the auditory stimuli. The looks towards the first NP align with the antecedent selection preference task; around 65% of the time the participants chose the first NP to be the antecedent of the pronoun. These suggest that it was indeed the case that the participants were resolving the pronoun. Furthermore, other work using a visual world paradigm has found effects right around the critical window when the anaphoric dependency needed to be resolved (e.g., J. E. Arnold et al., 2000; Clackson et al., 2011; Cozijn et al., 2011; Han et al., 2021; Kaiser et al., 2009; Runner & Head, 2014). Hence, it is hasty to conclude that the eyetracking visual world paradigm method is not sensitive to capturing real-time pronoun resolution.

This leads us to the second possibility, which relates to the type of construction that was tested. In most previous visual world paradigm studies that investigated the anaphoric dependency resolution, the critical pronoun was at the argument position, as exemplified in (9). In these studies, the primary eye-gaze data were observed around the pronoun window (around *him/her*(self)).

(9) a. The young boy was spending a day at the beach. He was amazed to see that the old man who was carrying a bucket built **him(self)** a magnificent sand castle. (Han et al., 2021)

b. The pharmacist that Molly met drove **her(self)** to the party. (Runner & Head, 2014)

c. Peter told Andrew about the picture of **him(self)** on the wall. (Kaiser et al., 2009)

In the current study, on the other hand, a possessive pronoun was the target, and this has a different argument status than *him/her(self)* used in earlier studies. Existing work has shown that the different argument statuses of the gaze target can lead to a different timing effect. For example, Koring, Mak, and Reuland (2012) compared the timing of reactivation of arguments when an unaccusative verb (e.g., *fall*) and an agentive verb (e.g., *jump*) were used using a visual world paradigm.[8] The authors found that with unaccusative verbs, the eye gaze towards the target argument increased at a delayed time point, after the verb offset. On the other hand, as for agentive verbs, the looks to the target argument were shifted at an early time, only slightly after the verb onset. This is a case in point that shows that the timing of eye-gaze shift can be influenced by the argument status of the target. Given this previous finding, it is possible that the delayed timing in the current visual world paradigm experiment, as well as the timing difference between the self-paced reading task and the eyetracking experiment, is due to the argument status of the possessive pronoun.

8. They also include a third type of verb—they refer to as a "mixed verb" (e.g., *sparkle*)—which is similar to an accusative verb in the type of thematic role that it assigned to the argument but also to an agentive verb in the syntactic status of the argument.

6.6.3 Limitations and future work

Discourse effect vs. temporal advantage

I suggested that the overlap in discourse structure information between the encoded linguistic representations leads to a competition effect. Yet, rather than the discourse structure effect, one could propose an alternative account that the ARC condition simply "benefited" from the extra time given before the RC. There was an additional time of 300 ms with the pause inserted before the onset of *who* in the ARC condition. As earlier studies show, a longer time to encode can make the linguistic representations stronger and more salient in memory (Karimi et al., 2020, a.o). Hence, the second NP always had the advantage of getting additional time to be encoded. This naturally could have led to higher activation in memory. If this were the case, it could be the longer encoding time rather than the discourse structure overlap that made the ARC condition exhibit a weaker competition effect. Given this possibility, a follow-up experiment could use filler words and disfluency such as "umm" or "well" (e.g., Lowder & Ferreira, 2016) instead of a pause before the RC to tease apart the effect of discourse structure information vs. extra time to encode.

Time sensitivity in a web-based visual world paradigm

As noted earlier, there can be delayed effects in the web-based eye-tracking experiment (Degen et al., 2021; Semmelmann & Weigelt, 2018; Slim & Hartsuiker, 2022; Vos et al., 2022). Binning the data into 50 ms for the analysis (Vos et al., 2022), or setting a calibration score threshold to filter out data points or participants (Lee & Kaiser, 2022) has been used as a way to remedy the uneven sampling rate. However, it is yet undetermined what the best practices are, specifically, how frequent calibration needs to take place during the experiment and what threshold score is most optimal. These all remain for future work.

Yet, even when the issue of calibration is addressed, given the nature of a web-based

experiment (in fact even in lab-based settings), gaps occur in collecting data points. As proposed in Tan, Kriegman, and Ahuja (2002) and Coe et al. (2022), for example, a possible way to address this issue could be to modify the data point interval by *linear interpolation* for the missing timestamps. While keeping the original data points, we can fit a linear interpolant between data points, expressed in X and Y coordinates. Given this interpolant "line," additional data points can be interpolated between, or extrapolated external to the existing data points along the line. Coe et al. (2022) apply this method with an interval of 2 ms and fill in the missing 169 timestamps. Future work can apply this technique to minimize the imbalance in the data point interval while being aware of the statistical assumptions when interpolation is applied to the data. It would also be useful to replicate the current data in a controlled laboratory setting and compare the results, as done in Vos et al. (2022).

6.7 Conclusion

In this Chapter, I used an intra-sentential anaphoric dependency that has two grammatically plausible antecedents for the pronoun, *their*. The two antecedents differed in their discourse status in two conditions. The RRC-embedding condition had both of the antecedents being part of the main clause; the ARC-embedding condition had the antecedents in distinct discourse structures, one in the main and the other in the subordinate discourse structure. Naturally, the two antecedents shared the same discourse-structure information in the RRC condition whereas in the ARC condition, they did not overlap in discourse status. The overlap in discourse structure information during encoding led to a competition effect, which was indexed by (smaller) eye-gaze differences between the two antecedents. The discourse-overlapping RRC condition had smaller eye-gaze differences compared to the non-overlapping ARC condition. Results from two web-based eyetracking visual world paradigm experiments suggest that this eye-gaze difference is manifested after the second antecedent is encoded. In

other words, the competition effect is realized fairly early, even before retrieval (for a subject at the matrix verb or an antecedent at the pronoun window) is triggered.

CHAPTER 7

GENERAL DISCUSSION

7.1 Summary of the experiments

The dissertation consisted of three sets of, a total of 8 experiments.

The first three experiments (Experiments 1–3), conducted in a self-paced reading task, made use of a standard number agreement attraction effect. The effect was used as an index to examine the retrieval interference effect induced by the linguistic entities inside the main and the subordinate discourse structure. The attraction effect was found in all three experiments in the baseline RRC condition. The ARC condition also showed the attraction effect in Experiments 2–3, but the effect was not found in Experiment 1. The key difference between Experiment 1 and Experiments 2–3 in the ARC condition was the active state of discourse questions at the point of retrieval. While the distractor in Experiments 2–3 was associated with the active questions at retrieval, the distractor in Experiment 1 was related to an inactive question.

The second set of experiments (Experiments 4–6), also conducted in a self-paced reading task, examined anaphoric resolution. There were two grammatically plausible antecedents for the pronoun (*their*). Crucially, the key contrast was made in whether the two antecedents were part of the same discourse structure or not. It was an empirical question whether such distinction would lead to the easiness/difficulty in pronoun resolution. Results showed that the reading times were longer when the possible antecedents were part of the same discourse structure (as in the RRC condition) compared to when they were in distinct discourse structures (as in the ARC condition). There were no reading time differences between the two RC conditions when the two possible antecedents were part of the same discourse structure. However, this contrast was no longer present when there was no anaphoric dependency (Experiment 6).

The third set of experiments (Experiments 7–8) used a visual world paradigm using an eyetracking method, with the same type of anaphoric dependency used in the second set of experiments. In these eyetracking experiments, the competition effect during the encoding stage of linguistic representations was examined. When the encoded linguistic representations shared similar information (as in the RRC condition), there were smaller eye-gaze differences between the target noun phrases compared to when they did not (as in the ARC condition). The eye-gaze differences between the two antecedents were more pronounced in the ARC (no-overlap) than in the RRC (overlap) condition. This encoding interference effect was at a fairly early stage, even before retrieval was triggered.

7.2 Findings and implications

I revisit the three questions initially raised at the beginning of the dissertation:

1. How does discourse structure information guide retrieval, given that discourse is dynamic and cannot be represented in a binary, static way?

2. How does discourse structure information guide retrieval, given that discourse information does not determine grammaticality?

3. How does discourse structure information have an effect on memory encoding?

7.2.1 Discourse question guides memory retrieval: an additional constraint

The first set of experiments addressed the first question. The number agreement attraction effect, an indication of retrieval interference, was observed in certain constructions in the ARC condition. The observation that even an ARC-embedded structure leads to a number agreement attraction effect suggests that it is insufficient to understand the retrieval process with a static, binary approach to discourse structure representation. If discourse structure

was represented as [+main] or [+subordinate], for example, and the memory retrieval mechanism uses these binary features as retrieval cues, the ARC structures would not have led to an attraction effect in any constructions. Yet, this was not the case.

Instead, the results suggested that the presence of the attraction effect was closely aligned with the active state of discourse questions at the point of retrieval. From this approach, subordinate (or secondary, less essential) discourse information can be just as readily accessible in working memory as the main (or primary, crucial) information when it is associated with an active discourse question. Even subordinate (ARC) content, can be a target for retrieval when the relevant discourse question is active. However, when the parser has completed processing the subordinate content and hence the discourse question related to the subordinate discourse is no longer active, the linguistic content associated with this inactive question no longer becomes the target for retrieval.

I have shown that this approach successfully accounts for the current experimental results as well as previous studies. The approach has the advantage that it offers an account on the way discourse structure information guides retrieval, even by capturing the dynamic aspect of discourse. It should be noted though that the discourse question-based approach is not an alternative framework to the cue-based retrieval theory. The core idea that linguistic dependencies are resolved under the content-addressable cue-based fashion is maintained, and the proposed question-based approach imposes an additional constraint on the cue-based retrieval framework.

A limitation of this discourse question-based approach is that it heavily relies on the way linguistic input and the questions are added to the discourse model. While the minimal units to be added to the *input* panel and the *question* panel were comparatively straightforward in the current experiments, the segmentation can be less clear-cut as the input sentences become more complex, specifically when they involve different types of discourse relations. The current work does not offer a guideline of how linguistic inputs can be chunked, and

what type of discourse questions can be generated associated with the linguistic inputs. This can be further addressed in future work by using different types of constructions and manipulating questions under discussion in the context, for example.

Furthermore, the discourse question-based approach would be most suitable for explaining the retrieval process involved in resolving a morphosyntactic agreement such as subject-verb agreement.[1] This is because of the nature of morphosyntactic dependency (subject-verb agreement in specific), which has a strict grammatical constraint on what can serve as an agreement controller (i.e., retrieval target). As briefly highlighted above, the question-based approach is not a standalone framework that makes predictions on what can be targeted and not for retrieval. A concrete framework on how memory retrieval operates is required, and only when a principle mechanism underlies the retrieval process can the question-based approach can be applied. The active status of a discourse question is an additional constraint that is imposed on memory retrieval.

7.2.2 When discourse structure information appears to be a retrieval cue

The first set of experiments has shown that discourse structure information, specifically the active discourse questions, plays the role of imposing an additional constraint on the retrieval target. Does this then mean that discourse structure information can potentially be used as retrieval cues?

The second set of experiments offers a response to this question and addresses the second research question. We have seen that in resolving the anaphoric dependency, the discourse status of the possible targets affected processing cost. Even when the preference for the antecedent—the antecedent in the same clause as the pronoun is preferred—was the same between the RRC and ARC conditions, the reading times to resolve the pronoun were dif-

1. It is possible to understand the pronoun resolution experiments from the question-based approach—a detailed description of the application of the question-based approach to pronoun resolution can be found in Appendix 7.3. However, I do not make strong claims about the question-based approach that it can be applied to dependencies where discourse structure does not decide the grammaticality.

ferent. The finding that the discourse contrast did not lead to the antecedent selection but affected the processing cost is noteworthy. It is a reflection that discourse structure information does not govern the grammaticality of an antecedent selection but affects the retrieval process involved in anaphoric dependency resolution.

Given this processing difference due to the discourse status of retrieval targets, can we say that discourse information serves as a retrieval cue? One could answer "yes" to this question if it is assumed that the inventory of retrieval cues includes any relevant information that influences and directs the process of *retrieval*. The observation that the discourse status of retrieval targets (i.e., antecedents) did not affect processing cost with the use of *the* stands useful for this argument. Since the retrieval process was initiated only in the *their*-condition but not in the *the*-condition, the finding suggests that discourse information was used for memory retrieval of the target. If the memory retrieval process is not just limited to understanding grammatically defined dependencies and not all retrieval cues need to participate in forming linguistic dependency, then discourse structure information can be considered as a retrieval cue.

However, if we accept discourse structure information as a retrieval cue, it needs to be defined how it can be represented. Note that we have already seen from the first set of experiments that a static, binary way of expressing discourse structure (e.g., [+main], [+subordinate]) does not hold. A way to address this is to express discourse structure information as an atomic state, such as [main] or [subordinate]. This is similar to marking salient or topicalized linguistic entities as [salient] or [accessible] (Kush, 2013; Kush, Johns, & Van Dyke, 2019) in resolving anaphoric dependencies. Insofar as salience or topichood plays a role in guiding and constraining anaphor resolution, discourse structure information serves as an influential factor in retrieving the antecedent, and we can adopt a similar way of expressing the discourse status.

The current work does not provide a formalized way of distinguishing types of retrieval

cues that are morphosyntactic-like from those that are discourse structure-like. The distinction can potentially be made by the setting gradient difference in the availability as a retrieval cue, or comparing weights of different retrieval cues can be made using cue-combinatorics (e.g., Van Dyke & McElree, 2011) through computational modeling. Or it can be a discrete one, where there is a strict boundary between a morphosyntactic cue vs. a discourse-related cue. These are all yet speculations. Further discussion on what qualifies a retrieval cue will be helpful for understanding processing linguistic dependencies that require non-discrete cues such as salience, prominence, accessibility (e.g., Kush, 2013; Kush et al., 2019; Parker et al., 2017).

7.2.3 Discourse-structure overlap during encoding leads to interference effect

The third set of experiments addresses the issue of encoding of discourse structure information, which pertains to the third research question. The findings collectively showed that the overlap in discourse structure information leads to an interference effect.

The current findings add to the scholarship on the encoding interference effect, which is comparatively minimal to the body of work on the retrieval interference effect. The second eyetracking experiment (Experiment 8) specifically showed that the interference effect due to the discourse-structure overlap could arise before retrieval is initiated. This finding contributes to the debate on whether an encoding interference effect is identified only when retrieval is triggered. While some have found that the encoding interference effect can be observed without retrieval of targets (e.g., Acheson & MacDonald, 2011; Rich & Wagers, 2020), others have shown that the effect is realized when retrieval is initiated (e.g., Parker & Konrad, 2020; Sekerina et al., 2016; Villata et al., 2018).

Why do we see a variation in the timing of when the encoding interference effect is realized? A possibility is that the timing depends on the type of linguistic information that leads to the interference effect. In fact, the early (i.e., before retrieval) vs. late (i.e., after/on

retrieval) encoding interference effect coincides with whether morphosyntactic features were used. For example, studies that found an early effect used linguistic information that does not involve morphosyntactic features but rather rhyme (1a), the type of discourse referent (1b) or semantic information (1c). In self-paced reading task studies, the overlap condition (examples in (i) in (1) took longer reading times than the no-overlap condition (examples in (ii). The effect was found at the second noun phrase region, or before the critical retrieval verb (*recently* in (1c)).

(1) Type of information used in studies that showed an early encoding interference effect

a. Rhyme (Acheson & MacDonald, 2011, Experiment 1)

(i) *The banker* that *the banker* sought bought the house.

(ii) *The runner* that *the banker* featured bought the house.

b. Referential form (Gordon et al., 2001, Experiment 3, subject-extracted RC)

(i) *The banker* that praised *the barber* climbed the mountain.

(ii) *The banker* that praised *Ben* climbed the mountain.

c. Semantic information (Rich & Wagers, 2020, Experiment 1)

(i) *The knife* that *the sword* was placed near had been recently sharpened.

(ii) *The knife* that *the shirt* was placed near had been recently sharpened.

On the other hand, studies showing a "late effect" used a morphosyntactic feature to control the overlap of the encoded linguistic representations. In (2), the grammatical gender feature overlap between the noun phrases was controlled, and the effect was realized at the retrieval site (around *surprised*) and not before.

(2) Morphosyntactic information used in studies showing late encoding interference effect (Villata et al., 2018, Experiment 1 (feminine head noun))

a. *La ballerina* che *la cameriera* ha sorpreso beveva un cocktail [...]
The dancer-FEM that the waiter-FEM has surprised drank a cocktail [...]

b. *La ballerina* che *il cameriere* ha sorpreso beveva un cocktail [...]
The dancer-FEM that the waiter-MASC has surprised drank a cocktail [...]
[Both] 'The dancer that the waiter has surprised drank a cocktail [...].'

Also in Parker and Konrad (2020), the subject-verb agreement (number agreement attraction effect) was examined. The encoding interference effect was found only beginning from the retrieval verb region, especially when the number feature was used for retrieval.[2]

This contrast may suggest that the type of linguistic information is tightly linked to the timing of when the encoding interference effect is realized. The discourse structure information used in the current study can be considered to have a similar status to non-morphosyntactic features such as the type of referential form, or semantic information. It is possible that morphosyntactic features are so strongly associated with grammatical dependency resolution that the parser uses those features as cues only when retrieval is triggered. Yet, this linkage remains to be further tested in future work. It will benefit from a meta-analysis of the current empirical findings on the timing of the encoding interference effect and the type of linguistic information that was used.

7.3 Conclusion

This dissertation examined the real-time comprehension of distinct discourse structure entities, those that are part of MAIN discourse structure and SUBORDINATE discourse structure.

2. Sekerina et al. (2016) used semantic plausibility information to examine the interference effect and found the effect beginning at the retrieval verb region ((i)). However, this is not a fair comparison to other studies above since the overlap of semantic information was only available at the verb. Only when the parser reaches the verb (e.g., *spotted*, *sewed*) can they target to retrieve the correct argument. Hence, before the retrieval site, there was no semantic overlap.

(i) Four images shown on a screen in a visual world paradigm
It was the {button / key / pen / earring} that the maid who returned from vacation {spotted/sewed} in the early morning.

The division of main vs. subordinate was made by using restrictive relative clauses (RRCs) and appositive relative clauses (ARCs)—RRC content was used to represent main, primary discourse structure information, and ARC content for subordinate, secondary information.

In order to understand how the working memory supports the real-time comprehension of discourse-structure information, two types of long-distance dependencies were used. The first set of experiments (Experiment 1–3, self-paced reading task) investigated the case of subject-verb dependency, specifically the number agreement attraction effect, with embedded ARC and RRC structures. The RRC structures exhibited the standard number agreement attraction effect in all three experiments. Meanwhile, the ARC structure showed the same attraction effect in Experiments 2–3 but not in Experiment 1. The next two sets of experiments (Experiments 4–8) used a construction involving an anaphoric dependency. Two plausible antecedents and the pronoun were included in the same sentence, while the antecedents differed in their discourse status. In the RRC condition, they were part of the same main discourse structure; in the ARC condition, one was part of the main while the other was part of the subordinate discourse structure. Experiments 4–6 (self-paced reading task) showed that when resolving the anaphoric dependency, longer reading times were spent when the two antecedents were part of the same discourse structure than when they were not. Results in Experiments 7–8 (visual world paradigm eyetracking method) suggested that encoding linguistic representations that share the same discourse-structure information (RRC condition) leads to an encoding interference effect. The eye-gaze ratio towards the two possible antecedents was compared, and the difference in the eye-gaze was bigger in the ARC condition, where the antecedents did not share the same discourse structure-related information.

The dissertation provides new empirical evidence that discourse-related information, even though it is not necessary for resolving linguistic dependencies, affects retrieval as well as the encoding process. Additionally, I have proposed to use a question-based approach to

understand the way discourse structure information guides retrieval. Incorporating the incremental update of discourse questions and their active status to the cue-based retrieval theory allows us to gain a comprehensive understanding of the current empirical findings.

APPENDIX

I demonstrate how the approach from an incremental update of discourse questions can potentially used to explain pronoun resolution. Here I focus on the structure used in Experiment 4, the key materials repeated below:

(3) a. Sentence-medial condition

The singers(,) who admired the violinists(,) invited their mentors to the party.

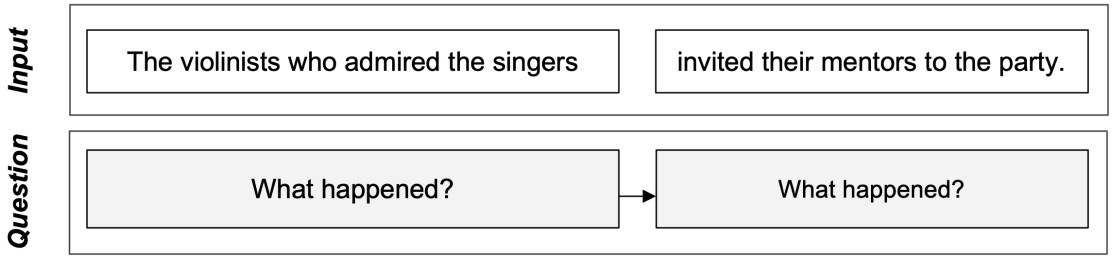
b. Sentence-final condition

The violinists admired the singers(,) who invited their mentors to the party.

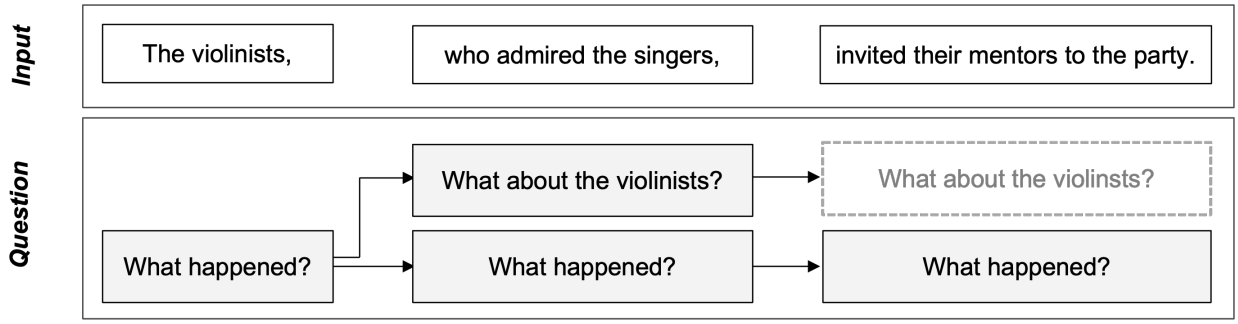
The empirical finding with the reading time results was that at the pronoun resolution site, the reading times were (i) longer RRC > ARC in the sentence-medial condition but (ii) not significantly different between the two in the sentence-final condition.

Figures 7.1 illustrate the incremental construction of discourse questions for the construction of the sentence-medial condition. For the sentence-medial RRC condition (Figure 7.1a), all discourse questions are active at the point of retrieval. By the time the parser processes the second input unit ("invited their mentors to the party"), the main discourse question, "What happened?" is still active (as the box marked in shaded grey). All linguistic units associated with this active question, including the two possible antecedents (i.e., *the violinists* and *the singers*), are active in the focus space. On the other hand, in Figure 7.1b, as the discourse progresses to the final linguistic input ("invited their mentors to the party"), the question regarding "What about the violinists" gets popped off. The linguistic content associated with this question about violinists contains information that *the violinists* admired the singers. As the discourse question gets popped off—no longer active (as shown in the box with dashed lines), the linguistic information regarding *the singers* is no longer in the focus space anymore. At the retrieval site of resolving the antecedent of the pronoun, the previous discourse question is inactive, and consequently, the second NP inside the ARC

subordinate content is not in the focus space.



(a) Sentence-medial RRC condition in Experiment 4



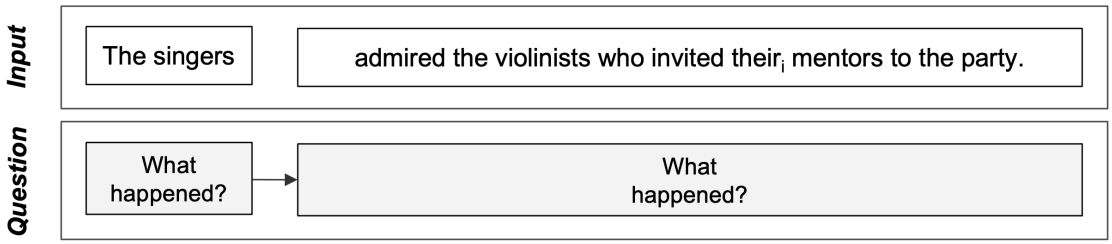
(b) Sentence-medial ARC condition in Experiment 4

Figure 7.1: Discourse questions in Experiment 4.
Note. The box with dotted lines indicates the discourse question is no longer active.

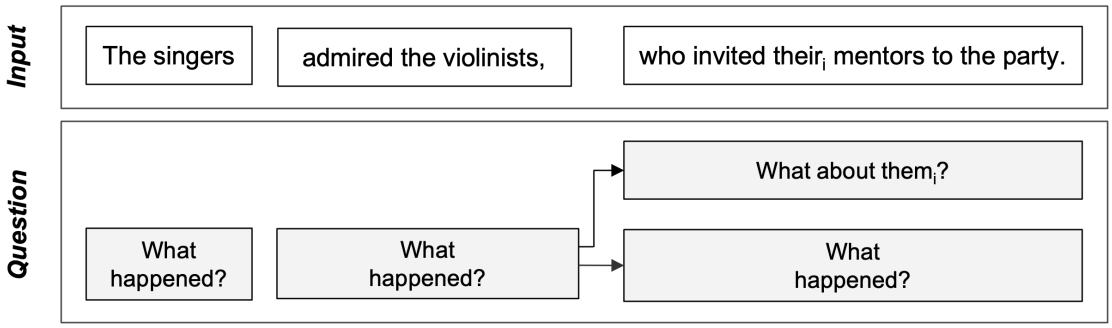
I highlight that the linguistic unit no longer being the focus of attention does not necessarily mean its semantic content or information is gone and not available in working memory. It is possible, as we saw in the number agreement attraction effect phenomenon (Chapter 4, that the grammatical, morphosyntactic features may not be available for retrieval. However, this does not mean that the information update that has been already made is removed out of the already-updated information state. In the sentence-medial ARC condition, the information that *the violinists admired the singers* is added to the common ground, updating the context set (as in AnderBois et al., 2015). Hence, even if the morphosyntactic features associated with the linguistic unit inside the ARC structure are no longer in the focus space, the discourse referents are available for anaphora.

The same approach can be applied to understanding the processing of the sentence-

final condition (Figures 7.2). Beginning with the RRC condition (Figure 7.2a), similar to other constructions, the RRC is in the same linguistic input as the head noun. All units raise the same discourse questions, and at the point of retrieval to resolve the pronoun, both antecedents are associated with the active discourse question. Next, the ARC adds a separate linguistic input and a subordinate discourse question is added to the question stack, "What about them?" While the question stack looks slightly different from that of the RRC condition, what matters is the active state of discourse questions at the point of retrieval. All discourse questions are active at the point of retrieval, similar to the RRC condition.



(a) Sentence-final RRC condition in Experiment 4



(b) Sentence-final ARC condition in Experiment 4

Figure 7.2: Discourse questions in Experiment 4.
Note. The box with dotted lines indicates the discourse question is no longer active.

I provide further details on how pronouns are resolved using the discourse question-based approach in pronoun resolution. The goal of using this representation is not to make claims about the semantic/pragmatic theory of anaphora resolution, but to show that this pronoun resolution process can take place independent of linguistic content being removed

from the focus space. I adopt the approach in the SDRT using the DRT-style of discourse representation. An illustration of how this theory is applied to model discourse was shown in Chapter 3.1.1 (specifically Figure **??**).

I use the example with the sentence-medial ARC structure, *the singers, who admired the violinists, invited their mentors to the party.* Each discourse unit for this given sentence can be labeled as in (4a). A well-formed discourse representation structure (DRS) can be formulated as in (4b).

(4) a. [The singers$_i$]π_1, [who admired the violinists$_j$,]π_3 [invited their$_{i/j}$ mentors to the party]π_2.

b. DRS

- $A = \{\pi_0, \pi_1, \pi_2, \pi_3, \pi_4\}$
- $\mathcal{F}(\pi_1) = K_{\pi_1}$
- $\mathcal{F}(\pi_2) = K_{\pi_2}$
- $\mathcal{F}(\pi_3) = K_{\pi_3}$
- $\mathcal{F}(\pi_0) = Coordination(\pi_4, \pi_2)$
- $\mathcal{F}(\pi_4) = Subordination(\pi_1, \pi_3)$

In the given example, the story unfolds by a discourse referent, *the singers*, is introduced to the discourse, followed by information that someone in the discourse *invited their mentors to the party. The singers* is in a *Subordinating* relation with the following unit, as expressed as $\mathcal{F}(\pi_4)$. This relation formulates π_3, and this chunk is *Coordinating* relation with the rest of the new content, π_4. This representation can be expressed with the DRT-style theory, as in Figure 7.3.

Now that we have a model of discourse structure, the theory about accessibility to previous units comes in to resolve the anaphora. Following the DRT-based definition of accessibility, a discourse referent u is *accessible* to an anaphoric discourse structure when the

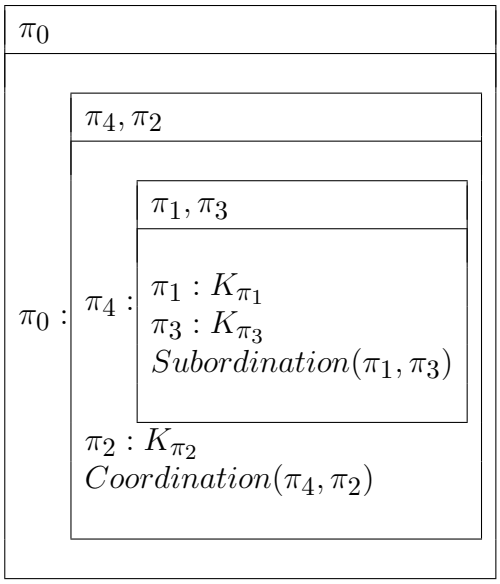


Figure 7.3: Discourse representation structure for anaphoric dependency in a boundary-crossing phenomenon.

structure that introduced the pronoun is equal or subordinate to the discourse structure that contains a possible referent. Given the subordinating relation that is identified in the exemplified discourse, a referent inside π_1 or π_3 is accessible to the pronoun in π_2. That is in our example, *the singers* in π_1 and *the violinists* in π_3 being accessible to *their* in π_3.

While re-anchoring to the prior referent for resolving the pronoun could be independent of the focus space, it can activate the encoded distinction in discourse structure information. Or, the other aspect of this is to say that if the re-anchoring does not happen, the discourse distinction is not as activated as in the other case. For example, if a definite article is used instead of a pronoun, re-anchoring would not necessarily take place, and the division would not be salient. This explains the findings in Experiment 6, where the division of discourse structure information was no longer present when *their* was replaced with *the*.

REFERENCES

Acheson, D. J., & MacDonald, M. C. (2011). The rhymes that the reader perused confused the meaning: Phonological effects during on-line sentence comprehension. *Journal of Memory and Language*, *65*(2), 193–207.

Alcocer, P., & Phillips, C. (2012). Using relational syntactic constraints in content-addressable memory architectures for sentence parsing. *Unpublished manuscript, University of Maryland, MD*. Retrieved from `http://www.colinphillips.net/wp-content/uploads/2014/08/alcocer_phillips2012_v2.pdf`

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Amaral, P., Roberts, C., & Smith, E. A. (2007). Review of the logic of conventional implicatures by Chris Potts. *Linguistics and Philosophy*, *30*(6), 707–749.

AnderBois, S., Brasoveanu, A., & Henderson, R. (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics*, *32*(1), 93–138.

Angele, B., Gutiérrez-Cordero, I., Perea, M., & Marcet, A. (2023). Reading(,) with and without commas. *Quarterly Journal of Experimental Psychology*, 1–11.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407.

Ariel, M. (1985). The discourse functions of given information. *Theoretical Linguistics*, *12*, 99–113.

Arnett, N., & Wagers, M. W. (2017). Subject encodings and retrieval interference. *Journal of Memory and Language*, *93*, 22–54.

Arnold, D. (2007). Non-restrictive relatives are not orphans. *Journal of Linguistics*, 271–309.

Arnold, J. E. (1998). *Reference form and discourse patterns* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*(2), 137–162.

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187–203.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The immediate use of gender information: Eyetracking evidence of the time-course of pronoun resolution. *Cognition*, *76*, B13–B26.

Asher, N. (1993). *Reference to abstract objects in English: A philosophical semantics for natural language metaphysics (Studies in Linguistics and Philosophy)*. Dordrecht: Kluwer.

Asher, N. (2000). Truth conditional discourse semantics for parentheticals. *Journal of Semantics*, *17*(1), 31–50.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge: Cambridge University Press.

Bach, K. (1999). The myth of conventional implicature. *Linguistics and Philosophy*, *22*, 327–366.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1–46). Dordrecht: Kluwer.

Bailey, H. R., Kurby, C. A., Sargent, J. Q., & Zacks, J. M. (2017). Attentional focus affects how events are segmented and updated in narrative reading. *Memory & Cognition*, *45*, 940–955.

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: doi:10.18637/jss.v067.i01

Beaver, D. I., Roberts, C., Simons, M., & Tonhauser, J. (2017). Questions under discussion: Where information structure meets projective content. *Annual Review of Linguistics*, *3*, 265–284.

Benatar, A., & Clifton, C., Jr. (2014). Newness, givenness and discourse updating: Evidence from eye movements. *Journal of Memory and Language*, *71*(1), 1–16.

Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition*, *25*, 653–660.

Birch, S. L., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language*, *34*(2), 232–267.

Birner, B., & Ward, G. (1994). Uniqueness, familiarity, and the definite article in English. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 20, pp. 93–102).

Birner, B. J., & Ward, G. L. (1998). *Information status and noncanonical word order in English* (Vol. 40). Philadelphia: John Benjamins.

Blutner, R., & Sommer, R. (1988). Sentence processing and lexical access: The influence of the focus-identifying task. *Journal of Memory and Language*, *27*(4), 359–367.

Boër, S. E., & Lycan, W. G. (1976). The myth of semantic presupposition. In *OSU Working Papers in Linguistics* (Vol. 21, pp. 1–90). Ohio State University.

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, *8*(2), 37–50.

Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, *108*(3), 896–904.

Brown-Schmidt, S., Byron, D. K., & Tanenhaus, M. K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, *53*(2), 292–313.

Bruce, K. B., & Farkas, D. F. (2007). Context structure for dialogues. *Unpublished manuscript, University of California, Santa Cruz*. Retrieved

from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=512ef8e75e4bbeb50bd43e367e10dc48a9ea5182

Büring, D. (2003). On d-trees, beans, and b-accents. *Linguistics and Philosophy*, *26*, 511–545.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: doi:10.18637/jss.v080.i01

Canac-Marquis, R., & Tremblay, M. (1998). The wh-feature and the syntax of restrictive and nonrestrictive relatives in French and English. In J. Lema & E. Treviño (Eds.), *Theoretical analyses on Romance languages* (pp. 127–141). Amsterdam: John Benjamins.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–94.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687.

Chen, Z., Jäger, L., & Vasishth, S. (2012). How structure-sensitive is the parser? Evidence from Mandarin Chinese. *Empirical approaches to linguistic theory: Studies of meaning and structure*, 43–62.

Christiansen, M. H., & Chater, N. (1994). Generalization and connectionist language learning. *Mind and Language*, *9*(3).

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*(2), 157–205.

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language*, *65*(2), 128–144.

Clackson, K., & Heyer, V. (2014). Reflexive anaphor resolution in spoken language comprehension: Structural constraints and beyond. *Frontiers in Psychology*, *5*, 904.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*(2), 259–294.

Clifton, C., Jr, & Frazier, L. (2012). Discourse integration guided by the 'question under discussion'. *Cognitive Psychology*, *65*(2), 352–379.

Clifton, C., Jr, & Frazier, L. (2018). Context effects in discourse: The question under discussion. *Discourse Processes*, *55*(2), 105–112.

Coe, B. C., Huang, J., Brien, D. C., White, B. J., Yep, R., & Munoz, D. P. (2022). Automated analysis pipeline for extracting saccade, pupil, and blink parameters using video-based eye tracking. *bioRxiv*, 1–29.

Colonna, S., Schimke, S., & Hemforth, B. (2015). Different effects of focus in intra-and inter-sentential pronoun resolution in German. *Language, Cognition and Neuroscience*, *30*(10), 1306–1325.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*.

Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. G. (2011). The time course of

the use of implicit causality information in the processing of pronouns: A visual world paradigm study. *Journal of Memory and Language*, *64*(4), 381–403.

Creemers, A., & Meyer, A. S. (2022). The processing of ambiguous pronominal reference is sensitive to depth of processing. *Glossa Psycholinguistics*, *1*(1).

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, *28*(1–2), 188–219.

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, *75*, 117–139.

Cunnings, I., & Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language*, *102*, 16–27.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49–59.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*(4), 465–480.

Dalmaijer, E. (2014). *Is the low-cost EyeTribe eye tracker any good for research?* (Tech. Rep.). PeerJ PrePrints.

Degen, J., Kursat, L., & Leigh, D. D. (2021). Seeing is believing: Testing an explicit linking assumption for visual world eye-tracking in psycholinguistics. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43).

Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, *40*(1), 172–201.

Dehé, N. (2014). *Parentheticals in spoken English: The syntax-prosody relation*. Cambridge: Cambridge University Press.

Dehé, N., & Kavalova, Y. (2007). Parentheticals: An introduction. In N. Dehè & Y. Kavalova (Eds.), *Parentheticals* (Vol. 106, pp. 1–24). Amsterdam: John Benjamins.

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351.

de Vries, M. (2006). The syntax of appositive relativization: On specifying coordination, false free relatives, and promotion. *Linguistic Inquiry*, *37*(2), 229–270.

de Vries, M. (2007). Parentheses as b-merged adverbial phrases. In *Parentheticals* (pp. 203–234). Amsterdam/Philadelphia: John Benjamins.

Dillon, B., Clifton, C., Jr, & Frazier, L. (2014). Pushed aside: Parentheticals, memory and processing. *Language, Cognition and Neuroscience*, *29*(4), 483–498.

Dillon, B., Clifton, C., Jr, Sloggett, S., & Frazier, L. (2017). Appositives and their aftermath: Interference depends on at-issue vs. not-at-issue status. *Journal of Memory and Language*, *96*, 93–109.

Dillon, B., Frazier, L., & Clifton, C., Jr. (2018). No longer an orphan: Evidence for appositive attachment from sentence comprehension. *Glossa*, *3*(1).

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory*

and Language, *69*(2), 85–103.

Drenhaus, H., Frisch, S., & Saddy, D. (2005). Processing negative polarity items: When negation comes through the backdoor. In S. Kepser & M. Reis (Eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives* (pp. 145–165). Berlin: Mouton de Gruyter.

Duff, J., Anand, P., Brasoveanu, A., & Rysling, A. (2023). Pragmatic representations and online comprehension: Lessons from direct discourse and causal adjuncts. *Glossa Psycholinguistics*, *2*(1).

Emonds, J. (1979). Appositive relatives have no properties. *Linguistic Inquiry*, *10*, 211–243.

Fabb, N. (1990). The difference between English restrictive and nonrestrictive relative clauses. *Journal of Linguistics*, *26*(1), 57–77.

Farkas, D. F., & Bruce, K. B. (2010). On reacting to assertions and polar questions. *Journal of Semantics*, *27*(1), 81–118.

Fedorenko, E., Babyonyshev, M., & Gibson, E. (2004). The nature of case interference in on-line sentence processing in Russian. In K. Moulton & M. Wolf (Eds.), *Proceedings of the 34 Annual Meeting of the North East Linguistic Society* (Vol. 34, pp. 215–226). Amherst, MA: GLSA.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, *27*, 285–319.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In B. Bel & I. Marlien (Eds.), *Proceedings of speech prosody* (pp. 83–90). Laboratoire Parole et Langage: SProSIG, Aix-en-Provence, France.

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

Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition*, *101*(1), 173–216.

Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249.

Frazier, L., & Clifton, C., Jr. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, *4*(2), 93–126.

Frazier, L., & Clifton, C., Jr. (2005). The syntax-discourse divide: Processing ellipsis. *Syntax*, *8*(2), 121–174.

Frazier, L., Dillon, B., & Clifton, C., Jr. (2018). Together they stand: Interpreting not-at-issue content. *Language and Speech*, *61*(2), 199–226.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

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. In A. P. Marantz, Y. Miyashita, & O. Wayne (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics*, *16*(2), 313–353.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407.

Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, *7*(1), 55–78.

Gildea, D., & Temperley, D. (2007). Optimizing grammars for minimum dependency length. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (pp. 184–191).

Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science*, *34*(2), 286–310.

Göbel, A. (2019). Final appositives at the right frontier: An experimental investigation of anaphoric potential. In M. T. Espinal, E. Castroviejo, M. Leonetti, L. McNally, & C. Real-Puigdollers (Eds.), *Proceedings of Sinn und Bedeutung 23* (Vol. 23, pp. 451–467). Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès).

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411.

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

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430.

Grant, M., Clifton, C., Jr, & Frazier, L. (2012). The role of non-actuality implicatures in processing elided constituents. *Journal of Memory and Language*, *66*(1), 326–343.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279.

Griffiths, J., & de Vries, M. (2013). The syntactic integration of appositives: Evidence from fragments and ellipsis. *Linguistic Inquiry*, *44*(2), 332–344.

Grillo, N., & Costa, J. (2014). A novel argument for the universality of parsing principles. *Cognition*, *133*(1), 156–187.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*(2), 261–290.

Grosz, B., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, *12*(3), 175–204.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274–307.

Gurevich, O., Johnson, M. A., & Goldberg, A. E. (2010). Incidental verbatim memory for language. *Language and Cognition*, *2*(1), 45–78.

Hale, J. (2001). A probabilistic Early parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)* (Vol. 2, pp. 1–8). Association for Computational Linguistics.

Han, C.-h., Moulton, K., Block, T., Gendron, H., & Nederveen, S. (2021). Pronouns are

as sensitive to structural constraints as reflexives in early processing: Evidence from visual world paradigm eye-tracking. *Frontiers in Psychology*, *12*, 611466.

Heim, I. R. (1982). *The semantics of definite and indefinite noun phrases* (Unpublished doctoral dissertation). Rutgers University, University of Massachusetts Amherst, MA.

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54*(3), 425–443.

Hobbs, J. R. (1985). *On the coherence and structure of discourse.* Stanford, CA: CSLI Technical Report 85-37.

Hofmeister, P., & Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology*, *5*, 1237.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* Oxford University Press: Oxford.

Huang, K.-J., & Dillon, B. (2023). *An infrequent, large cost underlies garden path effects: An RT distribution approach.* The 36th Annual Conference on Human Sentence Processing (HSP2023), University of Pittsburgh, PA. (March 9–11 (plenary talk))

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415.

Huang, Y. T., Spelke, E., & Snedeker, J. (2013). What exactly do numbers mean? *Language Learning and Development*, *9*(2), 105–129.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.

Hunter, J., & Asher, N. (2016). Shapes of conversation and at-issue content. In *Semantics and Linguistic Theory* (Vol. 26, pp. 1022–1042).

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, *55*(7), 3461–3493.

Jackendoff, R. (1977). *$\bar{X}$ syntax: A study of phrase structure.* Cambridge, MA: MIT Press.

Jäger, L. A., Benz, L., Roeser, J., Dillon, B. W., & Vasishth, S. (2015). Teasing apart retrieval and encoding interference in the processing of anaphors. *Frontiers in Psychology*, *6*, 506.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339.

Jasinskaja, K. (2016). Not at issue any more. *Unpublished manuscript, University of Cologne*.

Retrieved from https://dslc.phil-fak.uni-koeln.de/sites/dslc/katja_files/jasinskaja_any_more.pdf

Jun, S.-A., & Bishop, J. (2015). Priming implicit prosody: Prosodic boundaries and individual differences. *Language and Speech*, *58*(4), 459–473.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122—149.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201.

Kaiser, E. (2011). Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes*, *26*(10), 1625–1666.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, *112*(1), 55–80.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147.

Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, *23*(5), 709–748.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156.

Kamp, H. (1981). A theory of truth and semantic representation. In *Formal methods in the study of language* (p. 277-322). Amsterdam: Mathematisch Centrum.

Kamp, H., & Reyle, U. (1983). *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*. Dordrecht; Boston: Kluwer Academic.

Karimi, H., Diaz, M., & Wittenberg, E. (2020). Sheer time spent expecting or maintaining a representation facilitates subsequent retrieval during sentence processing. In S. Hartmann & U. Hahn (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society* (Vol. 42, pp. 2728–2734).

Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition*, *39*, 381–388.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Kehler, A. (2015). On QUD-based licensing of strict and sloppy ambiguities. In *Semantics and Linguistic Theory (SALT)* (pp. 512–532).

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602.

Koev, T. (2019). Parentheticality, assertion strength, and polarity. *Linguistics and Philosophy*, 1–28.

Koev, T. (2022). *Parenthetical meaning*. Oxford: Oxford University Press.

Koring, L., Mak, P., & Reuland, E. (2012). The time course of argument reactivation revealed: Using the visual world paradigm. *Cognition*, *123*(3), 361–379.

Kratzer, A., & Heim, I. (1998). *Semantics in generative grammar* (Vol. 1185). Oxford: Blackwell.

Kroll, M., & Wagers, M. (2019). Working memory resource allocation is not modulated by clausal discourse status. *Unpublished manuscript, University of California, Santa Cruz*. Retrieved from `https://people.ucsc.edu/~makroll/uploads/20190201_KrollLengthEffectsMs.pdf`

Kroll, M., & Wagers, M. W. (2017). Is working memory sensitive to discourse status? Experimental evidence from responsive appositives. In *XPrag 2017*. Cologne, Germany. Retrieved from `https://osf.io/view/xprag2017/`

Kush, D. (2013). *Respecting relations: Memory access and antecedent retrieval in incremental sentence processing* (Unpublished doctoral dissertation). University of Maryland, College Park, MD.

Kush, D., Johns, C. L., & Van Dyke, J. A. (2015). Identifying the role of phonology in sentence-level reading. *Journal of Memory and Language*, *79*, 18–29.

Kush, D., Johns, C. L., & Van Dyke, J. A. (2019). Prominence-sensitive pronoun resolution: New evidence from the speed-accuracy tradeoff procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(7), 1234.

Kush, D., Lidz, J., & Phillips, C. (2015). Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language*, *82*, 18–40.

Kwon, N., Ong, D., Chen, H., & Zhang, A. (2019). The role of animacy and structural information in relative clause attachment: Evidence from Chinese. *Frontiers in Psychology*, *10*, 1576.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, *82*, 133–149.

Lam, S.-Y., & Hwang, H. (2022). How does topicality affect the choice of referential form? Evidence from Mandarin. *Cognitive Science*, *46*(10), e13190.

Lascarides, A., & Asher, N. (2008). Segmented discourse representation theory: Dynamic semantics with discourse structure. In *Computing meaning* (Vol. 3, pp. 87–124). Dordrecht: Springer.

Lee, S. H.-Y., & Kaiser, E. (2022). Mapping language onto mental representations of object locations in transfer-of-possession events: A visual-world study using webcam-based eye-tracking. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, pp. 1270–1276).

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*, 339–359.

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

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

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159–191.

Liu, H., Xu, C., & Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, *21*, 171–193.

Logačev, P., & Vasishth, S. (2012). Case, word order and prominence: Interacting cues in language production and comprehension. In M. Lamers & P. de Swart (Eds.), *Case, word order and prominence* (Vol. 40, pp. 198–216). Dordrecht: Springer.

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*(6), 713–733.

Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies: Evidence from the visual-world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1400.

Lowder, M. W., & Gordon, P. C. (2021). Relative clause effects at the matrix verb depend on type of intervening material. *Cognitive Science*, *45*(9), e13039.

Madsen, J., Júlio, S. U., Gucik, P. J., Steinberg, R., & Parra, L. C. (2021). Synchronized eye movements predict test scores in online video education. *Proceedings of the National Academy of Sciences*, *118*(5), e2016980118.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1231.

Matin, E., Shao, K.-C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*, 372–380.

McCawley, J. D. (1982). Parentheticals and discontinuous constituent structure. *Linguistic Inquiry*, *13*(1), 91–106.

McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14*(6), 648–652.

McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, *43*(16), 1735–1751.

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

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 817.

McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation*, *46*, 155–200.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91.

McInnerney, A., & Atkinson, E. (2020). *Syntactically unintegrated parentheticals: Evidence from agreement attraction.* The 33rd Annual CUNY Human Sentence Processing, University of Massachusetts Amherst: Amherst, MA. (March 19–21 (oral presentation))

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking

looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*(5), 661–671.
Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), B25–B33.
Murray, S. E. (2014). Varieties of update. *Semantics and Pragmatics*, *7*(2), 1–53.
Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269.
Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5-6), 389–395.
Ng, A., & Husband, M. (2017). *Interference effects across the at-issue/not-at-issue divide: Agreement and NPI licensing.* The 30th Annual CUNY Human Sentence Processing, MIT: Cambridge, MA. (March 30–April 1 (poster presentation))
Nicenboim, B., Schad, D. J., & Vasishth, S. (2023). *Introduction to Bayesian data analysis for cognitive science.* `https://vasishth.github.io/bayescogsci/book/`. (Accessed: December 6, 2023)
Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language*, *99*, 1–34.
Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 5–19.
Niehorster, D. C., Cornelissen, T. H., Holmqvist, K., Hooge, I. T., & Hessels, R. S. (2018). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior Research Methods*, *50*, 213–227.
Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, *1118*(1), 155–167.
Nivre, J., De Marneffe, M.-C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., ... Zeman, D. (2020). Universal Dependencies v2: An evergrowing multilingual treebank collection. *arXiv preprint arXiv:2004.10643*.
Nouwen, R. (2007). On appositives and dynamic binding. *Research on Language and Computation*, *5*(1), 87–102.
Obata, M., Lewis, R. L., Epstein, S., Bartek, B., & Boland, J. (2010). Featural analysis and short-term memory retrieval in on-line parsing: Evidence for syntactic, but not phonological, similarity-based interference. In Y. Fainleib, N. LaCara, & Y. Park (Eds.), *Proceedings of NELS 41: Conference of the North East Linguistics* (Vol. 2, pp. 69–82). Amherst, MA: University of Massachusetts Graduate Student Linguistics Association.
Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition*, *37*, 346–357.
Oberauer, K., & Bialkova, S. (2009). Accessing information in working memory: Can the focus of attention grasp two elements at the same time? *Journal of Experimental Psychology: General*, *138*(1), 64.
Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758.
Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory.

Journal of Memory and Language, *55*(4), 601–626.

Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language*, *58*(3), 730–745.

Ooms, K., Dupont, L., Lapon, L., & Popelka, S. (2015). Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental setups. *Journal of Eye Movement Research*, *8*(1).

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 3839–3845).

Parker, D., & An, A. (2018). Not all phrases are equally attractive: Experimental evidence for selective agreement attraction effects. *Frontiers in Psychology*, *9*, 1566.

Parker, D., & Konrad, K. (2020). Teasing apart encoding and retrieval interference in sentence comprehension: Evidence from agreement attraction. In *Proceedings of the Annual Conference of the Cognitive Science Society* (Vol. 42).

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2017). The cue-based retrieval theory of sentence comprehension: New findings and new challenges. In T. Parodi, L. Escobar, & T. Vicenç (Eds.), *Language processing and disorders* (pp. 121–144). Newcastle: Cambridge Scholars Publishing.

Partee, B. (1975). Montague grammar and transformational grammar. *Linguistic Inquiry*, *6*(2), 203–300.

Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology*, *7*, 329.

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

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–428.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Ed.), *Experiments at the interference, syntax and semantics* (Vol. 37, pp. 153–186). Bingley, UK: Emerald Publications.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, *12*(5–6), 601–638.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633–654.

Potts, C. (2005). *The logic of conventional implicatures*. Oxford: Oxford University Press.

Potts, C. (2012). Conventional implicature and expressive content. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 3, pp. 2516–2536). Berlin: Mouton de Gruyter.

Prince, E. F. (1979). On the given-new distinction. In P. R. Clyne, W. F. Hanks, & C. L. Hofbauer (Eds.), *Fifteenth Regional Meeting of the Chicago Linguistic Society* (Vol. 15, pp. 267–278).

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, *17*, 133–140.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, *1*(2), 3.

Rich, S., & Wagers, M. W. (2020). *Semantic similarity and temporal contiguity in subject-verb dependency processing.* The 33rd Annual CUNY Human Sentence Processing, University of Massachusetts Amherst: Amherst, MA. (March 19–21 (oral presentation))

Riester, A. (2019). Constructing QUD trees. In M. Zimmermann, K. von Heusinger, & E. Onea (Eds.), *Questions in discourse* (Vol. 2, pp. 163–192). Leiden: Brill.

Roberts, C. (2004). Context in dynamic interpretation. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 197–220). Oxford: Blackwell.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*(6), 1–69.

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, *29*(8), 912–927.

Ross, J. R. (1967). *Constraints on variables in syntax* (Unpublished doctoral dissertation). MIT, Cambridge, MA.

Runner, J. T., & Head, K. D. (2014). What can visual world eye-tracking tell us about the binding theory. In C. Piñán (Ed.), *Empirical issues in syntax and semantics* (Vol. 10, pp. 269–286). Paris: CSSP—Colloque de Syntaxe et Sémantique à Paris.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: Evidence from eye movements. *Cognition*, *89*(1), B1–B13.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrase. *Cognitive Science*, *30*(2), 193–241.

Safir, K. (1986). Relative clauses in a theory of binding and levels. *Linguistic Inquiry*, *17*(4), 663–689.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163.

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1108.

Salverda, A. P., & Tanenhaus, M. K. (2018). The visual world paradigm. In A. M. B. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (pp. 89–110). John Wiley & Sons, NJ.

Sanford, A. J. (2002). Context, attention and depth of processing during interpretation. *Mind & Language*, *17*(1–2), 188–206.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*(9), 382–386.

Saslow, M. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Josa*, *57*(8), 1024–1029.

Schlenker, P. (2013). Supplements without bidimensionalism. *Unpublished manuscript, Institut Jean-Nicod and New York University*. Retrieved from `https://www.semanticsarchive.net/Archive/jgwMjNmM/Supplements_without_Bidimensionalism.pdf`

Sekerina, I. A., Campanelli, L., & Van Dyke, J. A. (2016). Using the visual world paradigm to study retrieval interference in spoken language comprehension. *Frontiers in Psychology*, *7*, 873.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*, 451–465.

Simons, M., Tonhauser, J., Beaver, D., & Roberts, C. (2010). What projects and why. In *Semantics and Linguistic Theory* (Vol. 20, pp. 309–327).

Slim, M. S., & Hartsuiker, R. J. (2022). Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIbex and WebGazer.js. *Behavior Research Methods*, 1–19.

Slioussar, N., & Malko, A. (2016). Gender agreement attraction in Russian: Production and comprehension evidence. *Frontiers in Psychology*, *7*, 1651.

Smith, M. E., Kurby, C. A., & Bailey, H. R. (2023). Events shape long-term memory for story information. *Discourse Processes*, *60*(2), 141–161.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299.

Sogo, H. (2013). GazeParser: An open-source and multiplatform library for low-cost eye tracking and analysis. *Behavior Research Methods*, *45*, 684–695.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447–481.

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics* (Vol. 9, pp. 315–322). New York: Academic Press.

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 550–569.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86.

Staub, A., Dillon, B., & Clifton, C., Jr. (2017). The matrix verb as a source of comprehension difficulty in object relative sentences. *Cognitive Science*, *41*, 1353–1376.

Sternberg, S. (1966). High-speed scanning in human memory. *science*, *153*(3736), 652–654.

Stewart, A. J., Holler, J., & Kidd, E. (2007). Shallow processing of ambiguous pronouns: Evidence for delay. *Quarterly Journal of Experimental Psychology*, *60*(12), 1680–1696.

Stine-Morrow, E. A., Shake, M. C., Miles, J. R., Lee, K., Gao, X., & McConkie, G. (2010). Pay now or pay later: Aging and the role of boundary salience in self-regulation of conceptual integration in sentence processing. *Psychology and Aging*, *25*(1), 168.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562.

Sun, C., & Breheny, R. (2020). Another look at the online processing of scalar inferences: An investigation of conflicting findings from visual-world eye-tracking studies. *Language, Cognition and Neuroscience*, *35*(8), 949–979.

Syrett, K., & Koev, T. (2015). Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics*, *32*(3), 525–577.

Tan, K.-H., Kriegman, D. J., & Ahuja, N. (2002). Appearance-based eye gaze estimation. In *Sixth IEEE Workshop on Applications of Computer Vision (WACV 2002) proceedings* (pp. 191–195).

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Temperley, D., & Gildea, D. (2018). Minimizing syntactic dependency lengths: Typological/cognitive universal? *Annual Review of Linguistics*, *4*, 67–80.

Thompson, C. K., & Choy, J. J. (2009). Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. *Journal of Psycholinguistic Research*, *38*, 255–283.

Truckenbrodt, H. (2015). Intonation phrases and speech acts. In M. Kluck, D. Ott, & M. de Vries (Eds.), *Parenthesis and ellipsis: Cross-linguistic and theoretical perspectives* (pp. 301–349). Berlin: Mouton de Gruyter.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134.

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., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316.

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

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

Vasishth, S. (2021). *A common mistake in psychology and psycholinguistics: Part 2.* `https://vasishth-statistics.blogspot.com/2021/08/a-common-mistake-in-psychology-and_13.html`. (Accessed: May 15, 2023)

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4), 685–712.

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models

of retrieval processes in sentence processing. *Trends in Cognitive Sciences*, *23*(11), 968–982.

Velleman, L., & Beaver, D. (2015). Question-based models of information structure. In C. Féry & S. Ishihara (Eds.), *The Oxford handbook of information structure.* Oxford: Oxford University Press.

Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, *9*, 2.

Voeten, C. C. (2022). buildmer: Stepwise elimination and term reordering for mixed-effects regression [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=buildmer` (R package version 2.6)

Voeten, C. C. (2023). permutes: Permutation tests for time series data [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=permutes` (R package version 2.8)

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the Visual World Paradigm. *Glossa Psycholinguistics*, *1*(1), 1–37. doi: doi:https://doi.org/10.5070/G6011131

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

Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology*, *67*(7), 1274–1304.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7-9), 905–945.

Ward, P., & Sturt, P. (2007). Linguistic focus and memory: An eye movement study. *Memory & Cognition*, *35*(1), 73–86.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112.

Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E–Z Reader. *Cognition*, *111*(1), 132–137.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 442.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*(6), 713–755.

Webber, B. L. (1981). Discourse model synthesis: Preliminaries to reference. In A. Joshi, B. H. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 283–299). Cambridge University Press.

Wilke, H. A. (2023). *The information structure of complex sentences: An empirical investigation into at-issueness* (Unpublished doctoral dissertation). The University of Edinburgh, Scotland.

Wolf, F., Gibson, E., & Desmet, T. (2004). Discourse coherence and pronoun resolution. *Language and Cognitive Processes*, *19*(6), 665–675.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency

types: ERP evidence. *Brain and Language*, *108*(1), 40–55.
Xiang, M., Wang, S., & Cui, Y. (2015). Constructing covert dependencies–the case of Mandarin wh-in-situ dependency. *Journal of Memory and Language*, *84*, 139–166.
Xu, P., Ehinger, K. A., Zhang, Y., Finkelstein, A., Kulkarni, S. R., & Xiao, J. (2015). TurkerGaze: Crowdsourcing saliency with webcam based eye tracking. *arXiv preprint arXiv:1504.06755*.
Yadav, H., Mittal, S., & Husain, S. (2022). A reappraisal of dependency length minimization as a linguistic universal. *Open Mind*, *6*, 147–168.
Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision making*, *16*(6), 1485–1505.
Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307.
Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273.
Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX).* doi: doi:10.17605/OSF.IO/MD832