UNIVERSITY OF CALIFORNIA
SANTA CRUZ

MEMORY KNOWS ITS BOUNDS: ENCODING CONTEXTS IN SENTENCE COMPREHENSION

A dissertation submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

LINGUISTICS

by

Lalitha Balachandran

September 2024

The Dissertation of Lalitha Balachandran is approved:

Professor Matt Wagers, Chair

Assistant Professor Hannah Hausman

Associate Professor Amanda Rysling

Professor Maziar Toosarvandani

Peter F. Biehl
Vice Provost and Dean of Graduate Studies

Copyright © by

Lalitha Balachandran

2024

Table of Contents

List of Figures

List of Tables

Author's Note

The experiments in Chapter 3 and Experiment 4 of Chapter 4 report joint work with Jack Duff (co-first author), Pranav Anand, & Amanda Rysling. Many thanks to undergraduate research assistants Jenna Maas, Claire Sakurauchi, and AJ Testerman for assistance with item creation and data collection for Experiments 3 and 4. Thanks also to Jenna Maas for assistance with item creation for Experiments 5 and 7.

For Maya.

Acknowledgments

I chose to come to Santa Cruz because I had a good feeling about the community. Since then, I've had the pleasure of working and laughing (and to be totally honest, stressing and crying) with so many wonderful people here, and in the process have learned a lot about friendship, collaboration, solidarity, and navigating life's challenges through a pandemic, natural disasters, several strikes, and the list goes on...but as I reflect on the past five years, I'm overcome with gratitude for the good memories I've shared with the people in my life during this time. I couldn't have done this without all of you.

I'd like to start by thanking my committee. Matt Wagers has had an obvious influence on the content of this dissertation, but he also took me under his wing and offered me guidance at a time when I doubted myself the most. He's taught me to be methodical and efficient, and to not stress the little things, but to make space for balance and joy, too. Perhaps most importantly, he allowed me a lot of freedom throughout this process and gave me the confidence to trust myself, because (much to my amazement) it always felt like he trusted me. This dissertation would not exist without him. I met Hannah Hausman later on in grad school, but I immediately knew I wanted to work with her. Her positive energy is infectious, and I have her to thank for making research feel fun again, even while racing to the finish line. I'm really grateful to her for being so willing to puzzle through the intricacies of language and data together. Thanks to Amanda Rysling who pushed me to be precise and steadfast, and whose influence is woven throughout this work. And thanks to Maziar Toosarvandani for always bringing exacting questions and challenging me to think critically and precisely – the formal sections of this dissertation especially benefited from his guidance.

I'm grateful to a number of other Santa Cruz linguists for their time and support throughout this journey. Pranav Anand wasn't an official member of my committee, but he might as well have been. He was a collaborator on Experiments 1-4, and the most supportive QP1 advisor I could have asked for. I'm grateful to him for pushing me to do better, but guiding me too, and offering me very practical advice in trying times. Thanks to Jaye Padgett and Ivy Sichel, and to Matt and Pranav, too, for their conversations and help developing EquiL, and for their commitment to equity in our department. I'm indebted to a number of other faculty, who I've been lucky to learn from, TA for, converse with, or receive a kind message from over the years: to Ryan Bennett, Adrian Brasoveanu, Junko Ito, & Armin Mester for their guidance during the foundational courses and beyond; to Sandy Chung, Donka Farkas, Jess Law, & Jim McCloskey for their consistent kindness and support over the years; to Jorge Hankamer for his detailed comments on my first QP; and to Roumi Pancheva for a fun seminar and her support during a challenging final quarter. Thanks also to our department's (current and former) staff, who keep things running so smoothly: Sarah Amador, James Funk, Ashley Hardisty, Logan Roberts, Gwyn Vandevere, and Maria Zimmer.

Without a doubt, I learned the most from and alongside my fellow grads. To my cohort – Vishal Arvindam, Myke Brinkerhoff, Dan Brodkin, Maya Wax Cavallaro, Yaqing Cao, and Allison Nguyen – thank you all for the camaraderie during the early years, and thanks especially to Vishal and Allison. Vishal has always reminded me that life and grad school are separate things, and has given me many a reason to enjoy a glass of wine, dance to groovy tunes, or sail away a Monday afternoon. Allison, thank you for your compassion and insight, the burrito picnics, sending me pictures of ridiculous/fabulous animals, and also for the grumpy crab who lived in my office and watched over me in the end stages of this dissertation.

I've been lucky to collaborate closely with several fellow grads over the years, who I'm grateful for as researchers and as humans. Thanks to Jack Duff for being a fellow

lover of music and discourse, an incisive co-author on portions of this dissertation and more, and for so patiently putting up with my chaos as a collaborator. Stephanie Rich – we started and ended as lab potatoes together, and I wouldn't have had it any other way. I've treasured our shared Aspirations and Chocolate Factory couch sessions. Some of the work I've been most proud of in grad school has been done side by side, through laughter and tears, with you. Thanks to Morwenna Hoeks, Nick Van Handel, and Cal Boye-Lynn for the countless hours spent on Nathaniel/corgis/kibble. Thanks also to Morwenna for the adventures, the conversations, and for loving liquid gold as much as I do; and to Nick, an excellent keeper, for taco bell and prosody, and for educating me on the eras.

Thanks to many elder linguistics grads, who I've relied on and looked up to over the years, for their friendship: especially to Steven Foley for so reliably checking in; to Margaret Kroll for the drinks, advice, and a shared love of sailing; to Jed Pizarro-Guevara for knowing how to make me laugh; to Kelsey Sasaki for the most fun conference reunions, featuring wine and hot tea; to Richard Bibbs for the board games, writing sessions, numerous good meals and late night hangs (and to Disco Richard for schooling everyone on the dance floor); and to Max Kaplan for bonding over shared experiences, both good and bad, and for one Very Adorable Cricket. Thanks to many other grads who came along later, and who I've celebrated and commiserated with over the years: to Mandy Cartner for shared perspectives on life, psycholinguistics, and cats; to Duygu Demiray for the laughs and the drinks; to Yağmur Kiper for honest conversations and impressive whistles; to Emily Knick for common hurdles to reflect on together; to Matthew Kogan for trivia and fun times; to Eli Sharf for the engaging conversations about all things psycholinguistics and beyond; to Sophia Stremel for the most amazing recipes and dinner parties; to Elifnur Ulusoy for their invaluable role in our teaching team during my first quarter as an instructor; to Niko Webster for always radiating positivity and for being methodical and determined in improving the grad life; and to all the other grads I've crossed paths with here, too.

I owe a big thanks to all the folks at UCSC's Teaching & Learning Center for their mentorship and role in my development as a teacher and education researcher: to Kendra Dority, Noori Chai, Roxanna Villalobos, Megan McNamara, Michael Tassio, and Herbie Lee. Thanks especially to Kendra Dority for her belief in me and for taking the time to support my goals, and to Megan McNamara for our joint writing sessions and her role in facilitating our EquiL workshops. I'm also grateful to be joining my postdoc advisors, Roxanne Beltran and Erika Zavaleta, in their labs very soon, and to be working with Lina Arcila Hernandez and Robin Dunkin, too. Thanks to all of them for giving me a reason to stay in Santa Cruz and to do work that I feel strongly about.

I never would have made it to grad school without the numerous mentors and peers who supported me prior to life in Santa Cruz. At UCLA, I owe thanks to Stephanie White and Madza Farias-Virgens for my first proper introduction to scientific research and for allowing me to be an honorary biologist for a while. I wouldn't have become a psycholinguist without Jesse Harris' willingness to entertain my ideas as an undergrad and his support in securing a job. Thanks to Colin Phillips for employing me for a very foundational year at UMD, and showing me what doing psycholinguistic research in grad school would be like. Thanks also to Phoebe Gaston and Hanna Muller for taking me under their wings when I was a lab manager, and to Ellen Lau for many intellectually interesting conversations and for being a source of support. Thanks to my fellow Baggett officemates: Aura Cruz Heredia, Zach Wellstood, and Anissa Zaitsu. Thanks especially to Zach and Anissa for the good times on both coasts: to Zach for being such a lovely person to spend time with, always picking up right where we left off, and to Anissa, who I immediately connected with and was pivotal in my decision to come to Santa Cruz. I'm grateful for all the weekends we've shared together here and in SF ever since. Thanks also to several others in the broader linguistics community who have devoted their time and expertise at various points: Mara Breen, Katy Carlson, Brian Dillon, Lyn Frazier, Ming Xiang, and many others.

I would not have made it through grad school without the non-academics in my life, who kept me tethered to reality and the outside world. Thanks to Georgia Schmidt and Goran Berg for being some of the first people who taught me how to practice new skills with consistency, and to follow what I love with persistence.

There's a small army of Santa Cruz sailors and boaters whose time and mentorship on the Monterey Bay kept me afloat while I was finishing this dissertation. Thanks a million to everyone at the UCSC boating center: to Amos Fishbein and Rusty Kingon for teaching me how to sail, to Tanya van Renesse for her advice and kindness on several occasions, and to Douglas Kisarale for teaching me to row. Thanks to Edith Lai and Ian Terry for all the work-and-chat sessions at Brady's. Thank you to all of the incredible sailors I met through Monday night Women's Sailing, and especially to Allison Jaballas (dockmaster extraordinaire), Lauren Fritz and Wesley Viebahn (for the sailing and movies and much more), Lauren Korth (for the pep talks and braving the foredeck together), and many others for fun times and peaceful evenings on the water. To Meghan, Kevin, and Welsey (again!) – thank you for helping me face the more stressful parts of boating. Finally, thanks to Todd Austin, Pat Barry, Gary Mirfield, and the Interlude/Sophie/Good Timin' crews for teaching me to race – fast is fun, indeed. I'm thankful to share a love for the ocean with all of you.

Thanks to Nicholas Boldt, Brennan Hines, and Ryann Powell, who have been such stable fixtures in my life: to Nicholas for keeping me on my toes, teaching me when to go with the flow and when to be spontaneous, and for being a loyal friend; to Brennan for 12+ years of both ordinary and memorable experiences together, and because I've loved growing into our friendship with such comfort and ease; to Ryann for our

shared taste in all the finer things in life (cats, cars, flowers, sauce, etc.), and for your willingness to do everything and nothing together. May we keep the extravehicular activities going until the end of our days. Thanks to Anjelica Casey, who's really a kindred spirit in so many ways, for the long chats and evenings spent together, for making Santa Cruz feel like home, and thank goodness for the day we bonded over baby goats; to Rachel Miller for being a fellow lover of animals and reality TV, for the occasional day trips out of Santa Cruz, and for quality hangs, whether in your backyard or at the disco; and to Dori Weiler for practical advice and being so fun and positive all the time. Thanks also to Uly Meheru, Ellen Fetherston, and Kellie Cheves for many years of unforgettable friendship.

I owe most of my determination (and stubbornness) in getting a PhD to my family – thanks to Jay and Kumar Balachandran for teaching me not to be a quitter and to question everything. And to Maya, the original potato, for her amazing resilience and being there through it all (-5 years). Love you immeasurably. I'm grateful to my beloved and mischievous cats, Noodle and Potato, for their silly and loving company always. And lastly, thanks to Nick Beber, my best friend and partner in every sense, for his unwavering love, patience, and support. I'm so lucky to live a life full of dreams with you.

Abstract

Memory Knows Its Bounds: Encoding Contexts in Sentence Comprehension

by

Lalitha Balachandran

Segmentation is a cornerstone of language processing across levels of linguistic analysis, and yet, standard models of linguistic memory leave the role of higher-order segments in online comprehension understudied. This dissertation advances the Context-Sensitive Encoding (CSE) hypothesis: that implicit prosodic boundaries (Bader, 1998; J. Fodor, 1998, 2002b) serve to partition sentences into distinct encoding contexts via a temporal context mechanism (Howard & Kahana, 2002) that shifts a gradually evolving contextual representation bound to item encodings at unambiguously marked prosodic boundary positions. In a series of reading and recognition memory studies, we demonstrate the role of CSE using three segmented sentence structures as test cases: appositive relative clauses, which have been shown to display *bypassing* of sentence-medial segments during online comprehension (Dillon et al., 2017; S. Kim & Xiang, 2022, 2023), and two types of focus-sensitive coordination that can prosodically separate their coordinates (*not only...but also* and *...as well as...*). The studies establish two consequences of CSE during sentence processing: previous contextual states may be reinstated at later points (a mechanism we term *Reinstantiation*), and in limited cases, the contents of a targeted segment may be accessed to the exclusion of other sentence content following cue-based retrieval (termed *Context-Sensitive Retrieval*). The account proposed here ultimately argues that bypassing stems from the interaction between (i) shifting the encoding context at prosodic boundary positions and (ii) anticipating upcoming subject-verb dependencies. We argue that this interaction can entirely account for an effect that has previously been attributed to idiosyncratic discourse properties of appositives.

Chapter 1

Introduction

It is a central property of human memory that we use temporal boundaries to partition our experience into uniquely identifiable episodes, which bring a wide variety of information along with them. In recalling a conversation with a friend, you may also recall information about the context in which the conversation was situated: memories about where you were (e.g., on a sailboat in the Monterey Bay), what the weather was like (moderately windy), and what time of day it was (right around sunset). Recalling any one of these contextual properties on its own may evoke memories of this particular conversation, and likewise, recalling the conversation may remind you of features of the context associated with it. Intuitively, spatiotemporal context is a powerful force in shaping and organizing human memories. In this example, it cannot be disentangled from the memory of a particular sentence uttered by your interlocutor; resurgence of such situational features is automatic and unavoidable.

Humans partition experience in an analogous manner at a much lower and more abstract level, too. This dissertation applies insights about how temporal context serves to organize words in list memory experiments to the domain of comprehending sentences. In particular, we appeal to a notion of *internal temporal context* that slowly evolves over time and reflects the state of the cognitive system at the time of encoding (Howard & Kahana, 2002). The dissertation will argue that such a temporal context mechanism allows the parser to mark significant structural positions, in order to manage storage of and access to linguistic content in memory at later points in time. Within psycholinguistics, two largely independent bodies of literature have investigated (i) how we search for word-level features in memory and (ii) the way in which linguistic segments serve as organizing units in memory. Here, we attempt to unite these literatures in order to make progress towards understanding how the partitioning of linguistic segments may affect memory retrieval processes at very short timescales: during the incremental comprehension of language, as a sentence unfolds word-by-word over time.

Modern models of linguistic memory rest on the premise that retrieval operations are content-addressable. Searching for a particular linguistic item within a sentence does not occur serially, from front-to-back or back-to-front, but rather proceeds rapidly on the basis of the available words in memory that were encoded with relevant features or *cues*. This framework has been termed *cue-based retrieval*, and is exemplified by the sentence in (36).

(1) Context: *Noodle and Potato are cats who love to hunt bugs.*
Noodle caught two **bugs**, and Potato caught three __.

Comprehenders reliably interpret this to mean that *Potato caught three bugs*, although the noun phrase *bugs* is elided from the second clause. Dependencies like this abound in language, and are resolved rapidly (within a few hundred milliseconds) without conscious effort on the part of the comprehender. Models of cue-based retrieval maintain that such dependencies involve retrieval of a previously encoded word from memory based on a search for features (like +NOUN PHRASE, for example), and that the speed of the this search operation proceeds in constant time (McElree et al., 2003). What may slow a comprehender down, however, is *cue overload*: that the number of competing items in memory encoded with similar features (e.g., the number of other items marked with +NOUN PHRASE) reduces the informativity of the retrieval cue, and thus gives rise to retrieval interference (Anderson & Bower, 1974). These findings lend crucial support to cue-based retrieval, which is well-equipped to explain reading and reaction time patterns representative of how we establish dependencies between non-adjacent words.

One consequence of adopting this influential approach in sentence processing has been that much of the modern psycholinguistics literature has focused on investigating item-to-item dependencies between non-adjacent elements, removed from the larger structure that a particular dependency is situated in. Despite this investigative trajectory, a myriad of evidence suggests that we attend to fine-grained structural distinctions during parsing, and that the memory representations of sentences as a whole encode information about structural boundaries across levels of linguistic analysis (e.g., representations of the syntax, prosody, and discourse). On the view that structural information of this nature is incrementally encoded alongside word-level properties, this naturally raises the question of whether and how structural domains may affect cue-based retrieval operations. Cases where larger structural domains do seem to matter have generated active areas of debate on the appropriate mechanistic approach. The issue of how to track syntactic relationships like C-COMMAND during comprehension is one such case that has received substantial attention, for example (see Kush, Lidz, and Phillips, 2015, for a discussion). These particular cases are not the focus of this dissertation, however. Instead, I turn my attention to a recent line of work on appositive relative clauses (ARCs), large sub-sentential units that align with syntactic, prosodic, and discourse boundaries. We start from the observation that these constructions display a curious processing phenomenon: across a variety of measures, sentences with ARCs, like the underlined portion of (2a), are easier to process than those with restrictive relative clauses (RRCs) like in (2b). That is, ARCs are perceived as less complex and their content does not interfere to the same extent as that of RRCs.

(2) a. The cat, who loves to hunt bugs, was chasing an elusive fly.

b. The cat that loves to hunt bugs was chasing an elusive fly.

Previous accounts of this phenomenon have attributed the behavior of appositives to their *not-at-issue* status – broadly, their semantic/pragmatic independence and irrelevance to the main point of the sentences that contain them. In a departure from these approaches, we claim that their processing ease is not due to particular discourse properties of ARCs, but rather stems from how the sentence-medial prosodic boundaries that surround them, cued by commas in text, interact with their representation in memory. Adopting an implicit view of prosody that assumes it can exert its influence during silent reading (Bader, 1998; J. Fodor, 1998), we garner evidence for this position based on a series of word-by-word reading and recognition memory experiments. We ultimately liken the processing of ARCs to other at-issue constructions that feature unambiguous prosodic isolation of sentence-medial units. Our results implicate a more general role for prosody in mediating segmentation in memory.

Against this backdrop, the dissertation will address the three central research questions in Table 1.1, using ARCs and the two structures in (3)-(4) as test cases, to probe the role of linguistic segmentation in memory during online sentence comprehension. Based on the results of Experiments 1-8, and on the basis that neither (3) nor (4) is interpreted in the "background" of the discourse[1], we claim that the effect of prosodic segmentation in memory plays a central role in (i) preserving linguistic memory for distinct segments and (ii) facilitating access to their contents. Furthermore, we contend that these effects can be captured via a temporal context mechanism, that not

[1]Throughout the dissertation, I will adopt the term *backgrounding* as a theory-neutral description of the sometimes peripheral discourse flavor of these and other parenthetical constructions.

Table 1.1: Central research questions of the dissertation.

❁ **Q1**: How does cue-based retrieval interact with the presence of linguistic segments in memory? Why can some segments be *bypassed*?
❁ **Q2**: Why do some prosodic boundaries appear to facilitate access to prior linguistic content, while others hinder access to prior content, e.g., through *Prosodic Visibility* (Schafer, 1997)?
❁ **Q3**: How does time spent on boundaries relate to memory and integrative processes for the contents of particular linguistic segments?
❁ **Q4**: Does accessing partitioned segments during retrieval have downstream consequences for interpretation?

only encodes linguistic boundary positions, but also allows the sentence processor to navigate and manipulate these segments during online comprehension.

(3) *not only* X*, but also* Y

Potato loves to hunt not only bugs, but also birds.

(4) X *as well as* Y

Noodle, as well as Potato, love to hunt bugs.

I will argue ultimately that looking beyond cue-based retrieval is a beneficial and necessary move to push psycholinguistic theorizing forward, as other corners of the memory literature are better-positioned to shed light on how the parser keeps track of higher-order structures in memory, and what the consequences of this may be for dependency resolution. To that end, we advance the *Context-Sensitive Encoding* (CSE) hypothesis in (5), which suggests that a representation of the temporal context modeled as a slowly changing vector randomly shifts at prosodic boundary positions (c.f. Howard & Kahana, 2002; Wagers, 2008), in order to create a partition in memory between previous and upcoming linguistic content. These partitions may then be used to facilitate later processing operations in several ways. Two of these are captured by Reinstantiation (6) and Context-Sensitive Retrieval (7), both processes that are reliant on a Context-Sensitive Encoding mechanism.

(5) **Context-Sensitive Encoding** (CSE) **Hypothesis**

Prosodic boundaries shift the *temporal context vector* (Howard & Kahana, 2002) during encoding, such that prosodic phrases partition syntactic content in memory into distinct *encoding contexts* (EC), and the content within an EC shares similar contextual features. These ECs are built up incrementally and may be reactivated on a by-context basis at a later stage, using positional cues. Segmented sentences therefore incur less processing load than non-segmented ones, due to decreased contextual interference during encoding and retrieval.

(6) **Reinstantiation** (to be revised): The contents of an earlier encoding context can be reinstated using positional cues in order to add additional material to a previously incomplete segment (c.f. Sederberg, Gershman, Polyn, & Norman, 2011).

(7) **Context-Sensitive Retrieval**: the cue-based retrieval mechanism can use the context vector as a cue, because retrieval of an item reinstates that item's encoding context. Therefore, retrieval of an item from a larger encoding context (i.e., one that contains more lexical/syntactic content) incurs greater contextual interference than that of a smaller encoding context.

Previous work suggests that participants are able partition experience within a linguistic task in a such a way that sentence-internal content produces interference during linguistic processing, but sentence-external linguistic content (like a word list) does not (Mertzen, Laurinavichyute, Dillon, Engbert, & Vasishth, 2024). This provides evidence that cue-based search processes can be restricted to the relevant linguistic context. The question remains whether the internal segments within a particular sentence themselves can serve to carve up the representation of a sentence in memory in a way that improves the precision of search processes. This is the focus here.

To preview our conclusions, we will show that encoding context shifts during sentence comprehension have particular utility for linguistic processing operations. First, the contextual state associated with a previous linguistic segment can be evoked again at the current timepoint, which mediates the accessibility of past content, and affects the encoding of subsequent content. This is the premise behind Reinstantiation (6). Second, we will show that Context-Sensitive Retrieval (7) plays a limited role, in that it serves to reduce interference following retrieval in segmented structures for some dependencies but not others, and that this may also have consequences for interpretive processes. The dissertation has another secondary goal: to disentangle the effects of *Prosodic Visibility*, which suggest that access to pre-boundary content is inhibited, from the observation that access to preceding content across appositive relative clause boundaries is facilitated. By the end, we aim to offer an account of when Context-Sensitive Retrieval and Reinstantiation take effect, and when Visibility takes hold.

1.1 Roadmap

Because the dissertation is intended for cross-discipline audiences and attempts to bring together work from several disparate areas, this section contains an outline of the contents of each chapter, along with a guide for how to approach the dissertation depending on whether the reader brings a perspective situated primarily in linguistics or in cognitive psychology. Throughout the following roadmap, I will indicate background sections that may be particularly relevant for linguists with a • symbol, and background sections that may be relevant for psychologists with ◦.

❧ **Chapter 2: The Bounds of Memory**

Chapter 2 first reviews evidence for segmentation in linguistic memory along divisions in syntactic (§2.1.1), prosodic (§2.1.2), and discourse (§2.1.3) structure, because these dimensions are relevant to our first test case, appositive relative clauses. The sections on prosody and discourse, in particular, describe common linguistic frameworks for understanding the structure and properties of each of these levels of linguistic analysis (○). The section on prosody also introduces the *Implicit Prosody Hypothesis* (Bader, 1998; J. Fodor, 1998), which underlies all experiments reported in the current dissertation (○). Section §2.2 contrasts online boundary effects in cases of Prosodic Visibility (Carlson et al., 2009; Schafer, 1997, i.a.) with those observed for ARCs. §2.3 introduces the memory frameworks utilized in the current work. §2.3.1 reviews the basics of cue-based retrieval; §2.3.2 explains the Temporal Context Model (Howard & Kahana, 2002) and motivates its relevance to word list and sentence memory (●); and §2.3.3 introduces the Context-Sensitive Encoding and related hypotheses (●, ○).

❧ **Chapter 3: Appositives in Memory**

Chapter 3 begins by introducing the relevant linguistic (§3.1; ○) and processing (§3.1.2; ●, ○) properties of appositives, while calling into question the conceptual and mechanistic feasibility of discourse-based accounts of ARC processing. In §3.1.3, it then spells out an account of how Context-Sensitive Encoding/Retrieval and Reinstantiation are assumed to operate in sentences containing ARCs, in contrast to accounts that appeal to the discourse status of these units (●, ○). Experiment 1 (§3.2) uses recognition memory profile of ARCs to motivate CSE; however, Experiments 2-3 (§3.3-3.4) do not find persuasive evidence for CSR in dependency resolution across ARC boundaries. A review of the previous literature suggests that Reinstantiation better accounts for the processing profile of ARCs.

❧ **Chapter 4: Coordinates in Memory**

Chapter 4 turns to an investigation of other segmented sentences, using the focus-sensitive coordinate structures in (3) and (4) as test cases. §4.1.1 first introduces the linguistic properties of *not only...but also* (○), then discusses their processing profile

(●, ○). Experiment 4 then contrasts the predictions of Context-Sensitive Retrieval and Visibility. §4.2 turns to a discussion of the linguistic properties of *as well as* (AWA) in §4.2.1 (○). Experiment 5 contrasts the predictions of Context-Sensitive Retrieval, Visibility, and Reinstantiation, ultimately finding evidence for the latter. §4.3 further establishes the relevance of Context-Sensitive Encoding and Reinstantiation for AWA coordinate structures by investigating how dwell time at their boundaries during reading relates to the memory representation of each of their segments. To that end, §4.3.1 discusses *clause-final wrap-up*, the process of dwelling at boundary positions during reading (●, ○). Experiment 6 then tests the predictions of Context-Sensitive Encoding and Reinstantiation, finding support for both.

❧ **Chapter 5: Reanalysis across Contexts**

Given evidence for CSE and Reinstantiation from Chapter 4, Chapter 5 searches for evidence of Context-Sensitive Retrieval in segmented structures that require syntactic reanalysis of multiple prosodically separated loci. §5.1 reviews the role of prosodic boundaries in constraining interpretation (○), per the *Intonational Domains Hypothesis* (Schafer, 1997). §5.2 (○) reviews the literature on syntactic reanalysis of NP/S ambiguities. §5.3 then details how the Context-Sensitive Retrieval, Reinstantiation, and Intonational Domains Hypotheses are predicted to interact with reanalysis processes (●, ○). Experiment 7 (§5.4) finds support for a limited role of Context-Sensitive Retrieval (i.e., in reanalysis but not other item-to-item dependencies). Experiment 8 (§5.5) investigates the resulting interpretations for NP/S structures and finds suggestive evidence that both Context-Sensitive Retrieval and Intonational Domains mediate interpretation.

❧ **Chapter 6** concludes by summarizing the main findings (§6.1), discussing future directions (§6.2), and taking stock of the theoretical implications (§6.3).

Chapter 2

The Bounds of Memory

2.1 Segmentation in Linguistic Memory

It is well-established that the capacity of short-term memory is tightly constrained. At any given time, finite resources are available for the active maintenance of sets of items in working memory, and very few items ($\sim$1-2) may occupy the focus of attention (i.e., be directly attended to) (McElree, 2006). Famously, these capacity limitations can be stretched by chunking groups of items in memory (G. Miller, 1956). Generally speaking, chunking a list into groups has been shown to improve memory for all of that list's content (Thalmann, Souza, & Oberauer, 2019). Whereas only the last item of a non-chunked list exhibits a benefit for being in the focus of attention, multiple items display this benefit in experimental settings where participants are encouraged to chunk items into groups (McElree, 1998). In this case, the items occupying the last chunk are privileged.

In non-linguistic list memory studies, groupings can be imposed along a number of dimensions of similarity between items: semantic similarity (Manning & Kahana, 2012), category membership (McElree, 1998), timing of presentation (Hitch, 1996), and more (see Kahana, 2012, for a review of relevant factors). Language is inherently different than lists, however, as sentences are composed of hierarchical structures across multiple levels of representation (e.g., the syntax, prosody, and discourse). As such, segmentation during sentence processing must occur at a variety of timescales, both small and large, across these levels. The resulting finer structural representation bestows a benefit, as memory for sentences is typically better than that of unstructured word lists (G. Miller & Isard, 1963). Although various dimensions of similarity matter in language processing, we assume that in sentence contexts, linguistic structure is an important guiding force in segmentation.

It has long been hypothesized that there is a relationship between the segmentation of large sub-sentential linguistic domains and units in memory. Research in this area has worked towards understanding which types of linguistic domains, and of what size, drive the formation of memory units. Just as the formation of non-linguistic chunks is flexible, so too is the formation of linguistic memory units. As suggested by Carroll and Tanenhaus (1978), either clause boundaries or certain boundaries below the clause level can trigger segmentation; that is, the likelihood of any given linguistic unit relating to a particular memory unit is not a property tied to specific boundary types. At the sentence-level, it's likely that a combination of factors determines the size of segmented units: the structure, length, and complexity of the sentence; individual differences in working memory capacity; in reading, individual differences in the *implicit prosody* assigned to a sentence (see §2.1.2.2); and others.

While segmentation is assumed to be a necessary process underlying the comprehension of language across timescales, chunking models have fallen out of favor in sentence processing since their introduction, because a substantial body of work has since evidenced that language processing is highly incremental (Marslen-Wilson, 1975). Linguistic structures are built in real-time, comprehenders have the ability to detect syntactic and semantic anomalies quickly, and they resolve long-distance item-to-item dependencies more or less immediately. This suggests that the language processor generally utilizes incoming information as soon as it becomes available. This incrementality is difficult to reconcile with models where processing is hypothesized to occur on a segment-by-segment basis, at clausal or prosodic boundaries (J. Fodor, Bever, Garrett, et al., 1974; Frazier & Fodor, 1978).

I concur with the argument that standard chunking models are too restrictive, both for the reason mentioned above regarding incrementality, and because they sometimes assume recoding of surface-level content into higher-order representations at segment boundaries. This has led to the paradoxical pair of assumptions that surface details are rapidly lost at major boundaries, but that at the same time syntactic features persist long enough to be utilizable for incremental processes across such boundaries (see C. Andrews, 2021, for an in depth related discussion). However, a subset of the psycholinguistics literature suggests that the presence of boundaries has a measurable effect on incremental sentence comprehension (Dillon et al., 2017; Kush, 2013; Schafer, 1997, i.a.), although researchers differ in their claims about what those effects are.

Ultimately, I will argue that insights from the domain-general memory literature on the role of temporal context during encoding (Howard & Kahana, 2002) can afford sentence processing theorists a model of both the incrementality needed to capture dependency resolution, as well as sensitivity to linguistic boundaries online. I provide a high-level overview of the proposal here, before introducing further background on linguistic segmentation. The account assumes incremental (i) encoding of word-level features in the usual manner (see §2.3.1), (ii) marking of significant boundary positions, and (iii) later sensitivity to the groups that emerge from process (ii), which I will term *encoding contexts*. The Temporal Context Model maintains that items are bound to a contextual representation during their encoding, which results in greater similarity in contextual features between temporally proximal items (or structurally proximal, in the case of sentences) than to temporally distant ones. Following Wagers (2008), I assume that certain linguistic boundaries induce a sudden "shift" in the contextual representation that renders subsequent input less contextually similar. Adjacent items spanning such a shift are hypothesized to be more featurally distinct from one another than the same two items would be in the absence of an intervening context shift. Thus, segment-level representations fall out naturally from the association between linguistic items and the contextual "markers" they are bound to, which encode information about their position in the larger syntactic structure they reside in, their temporal relations to nearby items, and other variable features of the internal and external context at the time of encoding. Crucially, the model need not assume that the contents of each encoding context is re-coded, although it can accommodate assignment of a context-level feature, if linguistically motivated (see §3.1.3). During comprehension, I propose that the processor is able to leverage contextual information (in a constrained way) so as to reduce potential interference from contextually dissimilar content. The details of this approach are spelled out in §2.3.3.

Recall from §1 that our starting point is the observation that appositive relative clauses (ARCs), like in (1), evidence sensitivity to structural boundaries online.

(1) The cat, <u>who loves to hunt bugs</u>, was chasing an elusive fly.

These constructions are unique because they involve alignment of significant syntactic, prosodic, and discourse boundaries. §3.1 discusses their properties in detail, but this chapter first reviews the empirical landscape of segmentation effects across each of these levels of analysis, because they have each been argued to correspond to units in memory. The present section discusses the role of segmentation in offline sentence memory studies, and the following section (§2.2) discusses online processing studies that have attested sensitivity to prosodic boundaries. In proposing a correspondence between linguistic segments and memory units, the particular types of linguistic boundaries that can induce the hypothesized context shifts is of central concern. To that end, the present section addresses two interrelated questions: (i) what counts as a "segment" in linguistic memory, and (ii) what types of segment boundaries are comprehenders sensitive to during incremental comprehension?

2.1.1 Syntactic Units

Early work in sentence processing explored the idea that syntactic clauses are the relevant unit of structure that serves to segment and organize linguistic content in memory (J. Fodor et al., 1974). G. Miller and Isard (1963) found that recall of sentences was significantly more accurate than that of lexically matched, randomly ordered lists, an observation termed the Sentence Superiority Effect. In a series of click dislocation experiments, J. A. Fodor and Bever (1965) found that listeners were more likely to veridically perceive the location of clicks when those clicks were positioned at clausal boundaries. When the true location of a click was clause-medial, it was "attracted" to the nearest boundary. That is, listeners displayed a tendency to perceive or remember clicks at nearby clause boundaries, regardless of their true position. Furthermore, this attraction did not apply to smaller constituent boundaries (Bever, Lackner, & Kirk, 1969). These findings lent support for the clause as a relevant segmentation and memory unit. However, a series of follow-up studies (Carroll & Tanenhaus, 1978; Tanenhaus & Carroll, 1975) determined that syntactic boundaries below the clause level can be sufficient to trigger segmentation. Specifically, they suggest the likelihood that any given unit will be segmented and therefore treated as a distinct memory unit is a not a fixed property of boundary type, but instead is dependent on functional completeness (i.e., whether a hypothetical segment contains a complete set of grammatical relations – subject, object, and verb) and the length of any given segment[2].

[2] A property that corresponds more straightforwardly with the phrasing of prosodic units; see §2.1.2.

Jarvella (1971) further advanced the idea that clause boundaries are used to segment syntactic content in memory. In a series of verbatim recall experiments, Jarvella (1971) investigated memory for sentences that contained a sequence of three clauses, where the medial clause was either incorporated into the second position of the first sentence, as in (2a), or into the beginning of the second sentence, as in (2b). In both cases, recall accuracy for the contents of the final clause was highest. Notably, recall accuracy for the medial clause was higher in (2b), where it was in the initial position of the second sentence, than in (2a), where it was in the final position of the first sentence. That is, memory for the medial clause was better when it began a sentence. On the basis of these results, Jarvella suggested that clauses must correspond to units in memory, but that only the most recently processed clause is a retrievable unit.

(2) a. $[_{S1}$ C1 C2$]$ $[_{S2}$ C3$]$

Kofach had been persuaded by the international to stack the meeting for McDonald. The union had even brought in insiders.

b. $[_{S1}$ C1$]$ $[_{S2}$ C2 C3$]$

The confidence of Kofach was not unfounded. To stack the meeting for McDonald, the union had even brought in outsiders.

It is important to note that these early proposals assumed that linguistic content was segmented before semantic interpretation occurred in a chunk-by-chunk manner. After the content of each clause was interpreted, it was suggested that its syntactic contents were lost in memory in order to free up working memory resources for the processing of the next clause (J. Fodor et al., 1974; Potter & Lombardi, 1990; Sachs, 1967). As mentioned previously, modern psycholinguistic theories have moved away from such an understanding of segmentation, because language processing occurs incrementally. But, more recent work highlights that structural domains matter for on online syntactic search processes.

For example, the resolution of dependencies such as reflexive anaphora and bound variable anaphora depends on the relational configuration between constituents in the syntactic strucutre, standardly defined in terms of a constraint on C-COMMAND[3] (Chomsky, 1981; Reinhart, 1983). In both cases, comprehenders obey this grammatical constraint online; for the most part, they do not consider antecedents located in inappropriate structural locations (Dillon et al., 2014; Dillon, Mishler, Sloggett, & Phillips, 2013; Kush, 2013), whereas other linguistic dependencies are notoriously prone to interference from structurally illicit distractors (Dillon et al., 2013; Wagers, Lau, & Phillips, 2009). The question of *how* such structural constraints influence search processes in memory has been a rich area of debate. Some theories of reflexive processing use the syntactic position of potential antecedents as a proxy structural cue (because C-COMMAND is a relational property, and cannot be a feature encoded on particular constituents). Of course, this type of sensitivity to structure is due to grammatical constraints on particular dependencies, and so it departs from the phenomena we investigate here. While such syntactic dependencies are not the focus of this dissertation, these cases have provided important insight into the fact that linguistic search processes in memory must have structural information at their disposal in some sense. Unfortunately, a full discussion of these examples would take me too far afield. I refer the interested reader to the references in this section for details about the mechanistic proposals, but discuss this issue no further here.

2.1.2 Prosodic Units

Recall Carroll and Tanenhaus's (1978) claims that (i) segmentation may occur below the clausal level and that (ii) the length of a syntactic phrase is a relevant factor in segmentation. Prosodic phrases offer a particularly apt place to search for a linguistic

[3]Node A c-commands node B if A does not dominate B, and the first node dominating A also dominates B.

structure-memory correspondence, because prosodic phrasing above the word-level (of phonological phrases, specifically) is flexible in that the position of a prosodic boundary can be sensitive to phrase length. In this section, I introduce two relevant pieces of background information before reviewing experimental evidence for a prosodic structure-memory correspondence. First, I discuss the Prosodic Hierarchy in §2.1.2.1, which provides a framework for analyzing the levels of prosodic structure present in sentences and utterances. I then introduce J. Fodor's (2002a; 2002b) Implicit Prosody Hypothesis (§2.1.2.2), which proposes that prosodic structure can exert its influence during silent reading, just as it does during listening. In §2.1.2.3, I summarize studies on overt and implicit prosody that lend support to the claim that prosodic boundaries affect the segmentation of linguistic units in memory.

2.1.2.1 The Prosodic Hierarchy

The Prosodic Hierarchy, a framework for representing the prosodic structure of sentences, is schematized in (3).

(3) The Prosodic Hierarchy

$$
\begin{array}{l}
u \ \text{= utterance} \\
| \\
\iota \ \text{= intonational phrase} \\
| \\
\varphi \ \text{= phonological phrase} \\
| \\
\omega \ \text{= prosodic word} \\
| \\
\ldots \ \text{= sub-word categories}
\end{array}
$$

Each level of the hierarchy corresponds to a different prosodic category, with smaller units (e.g., phonemes and syllables, below the word level) comprising the building blocks of larger ones (multi-word phrases). The current work deals only with prosodic units above the prosodic word level (ω), namely phonological phrases (φPs) and intonational phrases (ιPs), because these are the categories relevant to the prosodic segmentation effects considered here. Therefore, I set aside word- and subword-level phonological categories for the remainder of this section.

The φ-phrase is the level of the hierarchy directly above the prosodic word. φPs contain one or more words, and their existence is diagnosed by the presence of a Low (L-) or High (H-) phrasal tone at their right edge (Beckman & Pierrehumbert, 1986; Selkirk, 2000). Evidence for particular prosodic categories often comes from specific category-bounded phonological processes. One example of a process that diagnoses φ-phrases is the Rhythm Rule (4a), an optional rule which causes lexical stress to retract (i.e., to shift backwards by one syllable) in order to avoid a clash (two consecutively stressed syllables) within a single phonological phrase (Nespor & Vogel, 1986).

(4) a. Given the chance, rabbits ($_{\varphi}$réproduce quíckly) *reprodúce* $\rightarrow$ *réproduce*

b. Given the chance, ($_{\varphi}$rabbits reprodúce) ($_{\varphi}$véry quíckly) **réproduce*

(4a) exemplifies the optional application of the Rhythm Rule (RR) within a phonological phrase. In contrast, the RR does not apply to *reproduce* in (4b). This suggests that the RR does not perpetuate past phonological phrase boundaries.

The level of the Prosodic Hierarchy above the phonological phrase is the intonational phrase (ιP), which may consist of one or more phonological phrases. These domains typically bear a full intonational contour, and a single phonological utterance may contain several intonational phrases. Sentence-internal ιP boundaries are surrounded by prosodic breaks (perceptually significant pauses, which often align with syntactic boundaries) at their edges and accompanied by particular intonational correlates, like boundary tones, which may differ in realization depending on their discourse function. For instance, Pierrehumbert and Hirschberg (1990) suggest that the choice of boundary tone (L% or H%) indicates whether a particular ιP is forward-looking, indicating a relationship with the subsequent phrase. (5) presents one of their examples using the Tones and Break Indices (ToBI) annotation system for transcribing the intonational contours of English prosody (Beckman & Ayers, 1997). The first and second ιPs of (5) bear H% boundary tones at their right edge, indicating a forthcoming continuation and a relationship with the next phrase, whereas the final ιP bears an L% boundary tone, indicating the end of the utterance.

(5) ($_u$($_\iota$George likes cake L-H%)
($_\iota$He adores pie L-H%)
($_\iota$He'll eat anything that's sweet and calorific L-L%))

Sentence-internal ιP boundaries occur in specific types of syntactic constructions, including parentheticals (6), sentence-initial subordinate clauses (7), tag questions (8), and root clauses (9).

(6) ($_\iota$Potato,) ($_\iota$who's a tortoiseshell cat,) ($_\iota$loves to meow)

(7) ($_\iota$Whenever Potato meows,) ($_\iota$I give her a treat)

(8) ($_\iota$Noodle's a silly cat,) ($_\iota$isn't he?)

(9) a. ($_\iota$We call the neighborhood cat Ethel,) ($_\iota$but her name is actually Sadie)
b. We call ($_\varphi$the neighborhood cat Ethel,) ($_\varphi$and the alley cat Sadie)

In the examples (7), (8), and (9a), intonational phrases roughly correspond to clause-sized constituents in the syntax (compare (9a) to (9b), where sub-clausal constituents map onto φ-phrases). This is typical of intonational phrases; syntax-prosody mapping constraints, such as MatchClause, have been proposed to capture this correspondence (Selkirk, 2011). However, this mapping is violable, as is standardly assumed under constraint-based phonological frameworks (see McCarthy, 2011). Consider the first ιP of (6), which contains a single prosodic word, not a clause. In this sentence, which contains a medial appositive, the main clause is split into independent ιPs that do not comprise full clauses on their own: *Sadie* and *loves to meow*, respectively, in (6). Because ιP boundaries sometimes demarcate sub-clausal constituents, some accounts suggest that they map onto independent speech acts (see Ishihara, 2022, and §3.1), or units with distinct illocutionary force (e.g., assertion, question, exclamation, etc.) relating to speaker intent.

In general, units of syntactic structure do not always map directly onto units of prosodic structure; these cases are termed *syntax-prosody mismatches.* Two such mismatches are exemplified below, where syntactic constituents are denoted with square brackets and prosodic constituents are denoted with parentheses. (10) is a modified example from Simpson (2016), where the VP *weigh a few pounds* is interrupted by a φP boundary. (11) exhibits a case where an intonational phrase boundary and clause boundary misalign at the left edge of the appositive relative clause (*plants, which isn't actually...*). The prosodic properties of ARC syntax-prosody mismatches are discussed further in §3.1.1.3.

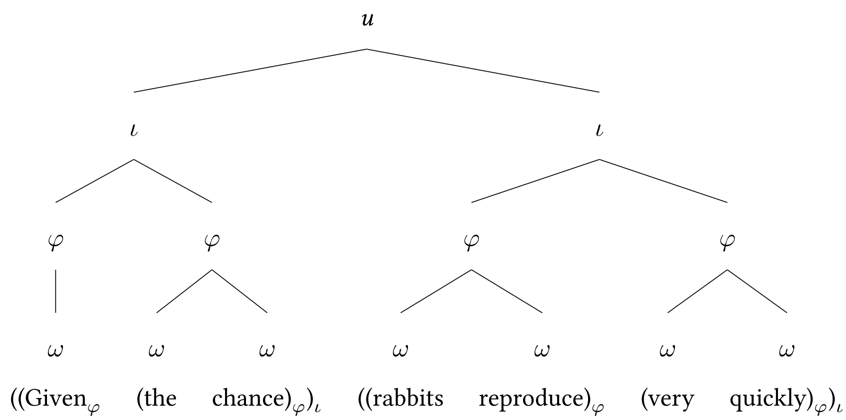
(10) ($_{\varphi}$Anyway) ($_{\varphi}$this cat must only [weigh like) ($_{\varphi}$a few pounds])

(11) ($_{\iota}$Potato likes to chew on plants [which) ($_{\iota}$isn't actually very good for her])

Such misalignments are especially abundant in naturalistic speech in the presence of fillers such as *like* in (10), or in cases where a parenthetical is particularly long. As mentioned previously, length (in number of syllables/words) is another factor that determines how much content may be phrased within a single φ-phrase. Similar mismatches are possible at all levels of prosodic structure, and part of the aim of phonological theorizing at the syntax-prosody interface is to account for when mismatches are possible. For the most part, we've seen that intonational phrase boundaries often coincide with clausal or discourse boundaries.

Equipped with a basic understanding of prosodic structure, we can now examine the structure for a sentence like (4b):

(12)



Throughout this discussion, I've left aside one important property of prosodic structure. A large body of theoretical literature supports the relevance of *recursivity* in the prosodic structure – the ability of a prosodic category to be contained within a category of the same type. Concretely, consider the difference in proposed structure for the following two coordinating constructions. For canonical list constructions as in (13) where each list item is relatively short, I assume an iterative prosodic structure, following Féry (2010) and others. This again exemplifies a case where the syntax-prosody mapping is non-isomorphic. In (13), each of the six φPs is immediately dominated by the ιP at the root of the prosodic structure.

(13) ((Anna made some errands and bought)$_\varphi$ (a bottle of orange juice)$_\varphi$, (an apple)$_\varphi$, (sugar)$_\varphi$, (butter)$_\varphi$, (and a pair of socks)$_\varphi$)$_\iota$

However, other coordinating constructions exhibit evidence of recursive φ-phrases (Féry, 2010; Wagner, 2010). The German example (14) from Féry (2010) exemplifies one such case.

(14) ($_\iota$ ($_\varphi$($_\varphi$Also jetzt steht) ($_\varphi$($_\varphi$links) ($_\varphi$der Gorilla))
↓($_\varphi$($_\varphi$rechts) ($_\varphi$neben dem Gorilla)) ($_\varphi$das Pferd)
↓($_\varphi$(und neben dem Pferd) ($_\varphi$rechts) ($_\varphi$der Löwe)))

'(So now the gorilla is standing to the left); (the horse is standing to the right beside the gorilla); (and the lion is standing beside the horse to the right)'

The evidence for the recursivity of (14) comes from the domain of *downstep*, a phonetic process where the pitch of a subsequent domain is lowered or compressed relative to each earlier one. Because this is a process that occurs across φ-phrases, relative pitch decrements between units of prosodic structure can be used to diagnose the presence of φ-phrase boundaries. On the other hand, pitch reset is typically taken to indicate the return to a higher level of prosodic structure following a series of prosodic phrases that are downstepped relative to one another. In (14), each line corresponds to the domain of downstep, and the ↓ symbol indicates overall pitch lowering relative to the previous φP. These domains correspond to dominating, maximal φPs in the prosodic structure, which each contain multiple embedded φPs.

In contrast to theoretical work, the convention in the prosodic sentence processing literature has been to assume iterative (non-recursive) structures. In the current work, I follow this convention for several reasons. At present, the sentence processing literature does not offer enough evidence to shed light on whether recursive prosodic structures are postulated by the parser on-line, or to what degree recursivity in prosodic structure is revised over the timecourse of processing a sentence. Following a large body of prosodic processing work on *implicit prosody*, the current studies consider the role of boundaries during silent reading. This is the focus of §2.1.2.2. In some ways, work on implicit prosody is in its infancy. Although it has been well-established that readers display sensitivity to prosodic structure online, there is still significant headway to be made in terms of understanding the types of prosodic information readers reliably posit in their mental representations. As such, it remains unclear whether phonetic processes that apply within particular phonological domains, such as downstep, are represented during reading. Moreover, we know little about the effect of recursive structures during listening. Any role of recursivity must first be established in listening studies before it can be investigated in the implicit prosody. Finally, it remains unclear to what degree the online segmentation mechanism relies on prosodic structures per se, as opposed to cues to pauses or temporal grouping more generally. I refer the interested reader to Van Handel (2022) for extensive discussion of other relevant considerations.

The linguistic constructions investigated in this dissertation (appositive relative clauses, *not only-but also* coordination, and *as well as*-coordination) each contain sentence-internal prosodic boundaries. The prosodic properties of these constructions are discussed throughout the remainder of the dissertation. For present purposes, I will assume that the presence of a boundary in these cases, whether recursive or non-recursive, will have essentially the same effect on incremental prosodic parsing, while acknowledging that this may very well be an overly simplistic approach. As we learn more about the relationship between recursivity and implicit prosodic representations, the field may be in a better position to hypothesize about its online effects.

2.1.2.2 The Implicit Prosody Hypothesis

Prosodic effects on sentence comprehension surface even during silent reading. Work on sentence memory shows that prosodic segmentation has an effect in listening and reading studies alike. Furthermore, all studies reported in this dissertation use silent reading and assume that readers postulate prosodic boundaries online, just as they do for other levels of linguistic structure. In order to motivate this claim, this section introduces an influential hypothesis that outlines a relationship between syntactic and prosodic parsing: the Implicit Prosody Hypothesis.

It has long been acknowledged, both by work in psychology and through general intuition, that silent reading involves some experience of hearing one's "inner voice". Crucially, this experience is connected in important ways to sentence memory and comprehension. Reading typically involves *subvocalization*, or inaudible articulation of speech likely associated with implicitly assigning a prosodic structure to a sentence. Early work by Slowiaczek and Clifton (1980) suggested that reading comprehension is degraded when this process is disrupted. The authors instructed participants to repeat nonsense syllables out loud while reading, and found that this resulted in worse understanding of the passages they presented than when the same disruption was present during a listening task. Memory for the passages was not significantly impaired in either task. They suggested that interrupting the process of assigning an implicit prosodic structure to a written passage drove the decrements in comprehension during reading.

Later investigations determined that implicit prosody also has consequences for parsing sentences in real-time. Beginning with Bader (1998) and J. Fodor (1998), it was proposed that the inner voice guides syntactic ambiguity resolution. In Bader's case, this explanation was put forth to explain why certain syntactic garden paths incur more processing difficulty and are more difficult to recover from (e.g., *The horse raced past the barn fell*). In other words, Bader sought to understand why certain types of syntactic reanalysis rise to the level of consciousness while others do not. In the most difficult of cases – the ones that comprehenders tend to become consciously aware of – syntactic reanalysis may never occur successfully. Bader suggested that such unrecoverable cases are more likely to result from "double" misanalyses, or those that require reanalysis of both syntactic and prosodic boundaries.

J. Fodor (2002a; 2002b) applied the same intuition to temporarily ambiguous sentences that have different cross-linguistic parsing and interpretive preferences. The relationship between prosodic boundaries and syntactic parsing preferences across a variety of constructions and languages led J. Fodor (2002b) to propose the Implicit Prosody Hypothesis (IPH) in (15).

(15) **Implicit Prosody Hypothesis**: (J. Fodor, 2002b)

In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for the construction.

Consider the following relative clause attachment ambiguity.

(16) a. Someone saw [$_{NP1}$the daughter] of [$_{NP2}$the actress] [$_{RC}$who was on the balcony].

b. Someone saw the daughter of the actress | who was on the balcony.

(16a) exemplifies an attachment ambiguity where the relative clause *who was on the balcony* may modify either NP1 *the daughter* or NP2 *the actress* of the complex object (*the daughter of the actress*). The default preference for (16a) is for the RC to attach low, i.e. to modify NP2. In the presence of a prosodic break, indicated via the '|' in (16b), there is a increased tendency for the RC to modify NP1 relative to (16a). This effect also holds in the implicit prosody, where the presence of such breaks is manipulated by commas other orthographic cues to a boundary, or by the length/prosodic weight of constituents (Hemforth et al., 2015; Jun & Bishop, 2014; Van Handel, 2022). These results suggest that prosodic breaks, whether implicit or overt, influence the syntactic parse and the resulting interpretation of a sentence. Furthermore, it has been observed cross-linguistically in silent reading that simply adding length to one of the constituents can override the default attachment preference, leading to a higher proportion of high attachment in the presence of a longer NP2, because a prosodic break is more likely to follow a heavy NP (Hemforth et al., 2015). Such length-driven preferences have led researchers to argue that syntactic parses that lend themselves to balanced, or roughly equal-sized prosodic phrases, are ultimately preferred. It should be noted that other studies (Hwang, Lieberman, Goad, & White, 2011; Van Handel, 2022) have found mixed results; see §2.2.2 for more detail.

Another well-established case study involves the incremental interpretation of NP/Z garden path sentences (17), which suggests that the prosodic parse of a sentence is posited incrementally and thus affects interpretations in a more or less immediate manner, rather than at a delay[4]. That is, the assignment of both implicit and explicit prosodic structure may constrain syntactic parsing predictions in advance (either predictively, or as soon as other guiding information is made available), suggesting that the effect of prosodic boundaries is not a post-hoc one that applies following syntactic parsing (Slowiaczek, 1981; Speer, Kjelgaard, & Dobroth, 1996), contra early views that privileged the syntax over incremental structure-building at other levels of representation.

(17) a. ($_\iota$While Anna dressed the baby) ($_\iota$ . . .)

b. ($_\iota$While Anna dressed the baby) ($_\iota$spit up! . . .)

c. ($_\iota$While Anna dressed)$_\iota$ ($_\iota$the baby spit up on the bed)

During on-line processing, given only the input in (17a), there is an initial, but ultimately incorrect, preference to parse the NP *the baby* as the object of the subordinate clause verb *dressed.* This is often attributed to a general syntactic parsing preference termed Late Closure: associate incoming lexical input with the phrase/clause

[4]Note that in the RC attachment cases, re-evaluating the prosodic parse based on the total number of words in a particular phrase cannot occur immediately, and so this must be done at some later stage.

currently being processed (i.e. attach low), as opposed to attaching it at a higher level of syntactic structure (Frazier & Fodor, 1978). When the main verb is encountered in (17b), this triggers reanalysis, as the NP *the baby* must be reinterpreted as the subject of the main clause, rather than the object of the subordinate clause (see §5.2 for a detailed discussion of reanalysis). Crucially, this particular example requires reanalysis on two levels of representation: (i) the prosodic structure, as the ι-phrase boundary at the right edge of the subordinate clause must be adjusted, and (ii) the syntactic structure, as *the baby* must be delinked from the subordinate clause verb and reattached as the subject of the main clause.

Such cases offer support to Bader's (1998) claim that double reanalyses are typically consciously detected by comprehenders and are generally more difficult to recover from. Successful reanalysis of (17b) requires arriving at the representation in (17c). But even when reanalysis is "successful", comprehension question responses suggest that both the globally grammatical interpretation and the initially pursued but incorrect one concurrently linger to some extent. That is, *Anna dressed the baby* and *the baby spit up on the bed* are both accepted as true in sentence-final comprehension question responses (Ferreira, Christianson, & Hollingworth, 2001). Thus, NP/Z garden paths constitute cases where initial prosodic misanalysis may lead to a globally ungrammatical interpretation. Another similar case, the so-called *prosodic local coherence* effect (Frazier, Clifton, Carlson, & Harris, 2014), is discussed in §5.1.

(16) and (17) then provide evidence in favor of the idea that the position of prosodic boundaries influences incremental syntactic attachment decisions and interpretations, and crucially, that this occurs even in the absence of an overt prosodic signal. Furthermore, expectations about the implicit prosody can be generated in advance and affect parsing. For example, in highly constraining metrical contexts (e.g., in limericks), the disruption of a repeating metrical pattern also results in a disruption in

reading (Breen & Clifton, 2011). Given a strong expectation for a recurring strong-weak-strong-weak stress pattern, a strong-weak-strong-strong sequence of syllables causes a slowdown in reading times, suggesting that the implicit prosody assigned to a portion of a sentence can serve to generate predictions about the prosodic status of upcoming material.

Because the current studies focus on the role of prosodic boundaries and how these boundaries relate to partitions in memory, the remainder of this section reviews previous work that speaks directly to boundary effects and their interaction with working memory constraints. Reading studies suggest that readers attend to orthographic cues, such as commas, and use these as evidence for the location of prosodic boundaries during real-time comprehension. As mentioned previously, attachment ambiguity studies have shown that the presence of a comma preceding a relative clause results in a preference for high attachment interpretations. Additionally, Steinhauer and Friederici (2001) find common event-related potential (ERP) responses to intonational phrase boundaries during listening and comma-marked boundaries during reading in that both elicit an ERP component called a *closure positive shift* (CPS), a positive-going waveform that is typically smaller in magnitude and occurs earlier than the P600 component, which is associated with the detection of syntactic anomalies and reanalysis. The magnitude of the CPS was smaller for visual compared to auditory stimuli, and was modulated by individual differences in comma use by participants in writing. This suggests two things: (i) that common cognitive processes are operative in both the visual and auditory processing of boundaries, but also (ii) that neural correlates of implicit prosody are subject to individual differences.

In the context of investigating clause- and sentence-final *wrap-up effects* (Just & Carpenter, 1980), Hirotani, Frazier, and Rayner (2006) investigate the Dwell Time Hypothesis, which proposes that readers dwell at clause and sentence boundaries because at these points during processing, final interpretive processes occur, and information is shunted out of working memory. Across two eye-tracking studies, Hirotani et al. (2006) found that readers tend to dwell longer on sentence-medial, clause-final words marked with a comma compared to those that lacked an explicit orthographic cue to a prosodic boundary. Though readers tended to dwell longer comma-marked words, overall reading times for sentences with medial punctuation were faster than for comparable sentences without punctuation. This was driven by the fact that post-comma reading times were facilitated. The authors ultimately argue that commas induce stronger wrap-up effects because they unambiguously signal the prosodic boundaries of a sentence, which delineate specific junctures at which interpretive wrap-up processes may occur and that these processes ultimately facilitate reading[5].

The studies reported in this dissertation (in part) attempt to manipulate the prosodic structure of experimental items through the use of commas. Adding commas in the orthography is not a choice without other linguistic consequences, as changes in prosodic phrasing often go hand in hand with changes in syntactic structure or discourse status (see §3.1.1.4 for more detail). The studies reviewed in this section emphasize that while introducing commas can be an effective way to induce a particular prosodic structure, it is not a foolproof one, because implicit prosodic phrasing is subject to considerable individual variation (Steinhauer & Friederici, 2001; Swets, Desmet, Hambrick, & Ferreira, 2007). Furthermore, not all readers use orthographic cues in an equally informative way (Steinhauer & Friederici, 2001). That is, they may not reliably insert a prosodic break in their implicit representation upon encountering a comma, and when they do, the specific prosodic nature of this break (e.g., its duration, surrounding boundary tones, length of the preceding syllable, etc.) is bound to vary across individuals, because these features vary in speakers' explicit prosody as well.

[5] There are a number of other explanations that have been put forward to explain wrap-up effects, which are a rather mysterious phenomenon in sentence processing. Sources of wrap-up are discussed further in §4.3.1, as this is the focus of Experiment 6.

In sum, phonological phrases in silent reading are noisier, as they are subject to more variation; this is discussed further in the context of Experiment 4 (§4.1), which uses a construction with medial phonological phrase boundaries. The issue of variability is less pressing for intonational phrase boundaries, as these almost always correspond to significant syntactic boundaries as well. Experiments 1-3 and 5-8 use commas to demarcate intonational phrase boundaries.

In the current work, we assume that commas act as meaningful cues to prosodically group syntactic content together. This assumption is partially supported by the fact that previous attempts at influencing prosodic structure via orthographic cues have been successful. Experiment 6 in particular validates the effectiveness of this choice for the particular constructions utilized here. The question remains whether these prosodic groupings have any effect on memory representations and later access to those representations. The prosodic assumptions of the experiments reported here stand on firmer ground because they contain sentence-internal intonational phrase boundaries with concurrent differences in syntactic and (sometimes) discourse structure. The corresponding limitation of these studies is that it is difficult to attribute any effects to the implicit prosodic representation alone, as opposed to other properties of their linguistic structure. As with all work that attempts to reason about implicit prosody, the experiments reported here should ultimately be replicated using overt prosody in order to draw firmer conclusions about the degree to which the reported effects are attributable to the prosodic structure, using the procedure proposed by J. Fodor (2002b) for establishing an implicit prosodic effect (18). Although the procedure in (18) makes specific reference to ambiguity resolution preferences, I assume that the implicit prosody has other effects on processing as well.

(18) **Procedure for testing the IPH**: (J. Fodor, 2002b)

a. Find a factor F which can be manipulated in an experiment, and which measurably affects the OVERT prosody of a sentence.

b. Show that the overt prosodic difference caused by F measurably influences an ambiguity resolution preference in parsing.

c. Show (or claim?) that F does not affect parsing DIRECTLY.

d. Include F in a silent reading task. Is ambiguity resolution affected by F as it is the listening task?

2.1.2.3 The Prosody-Memory Relationship

Prosody has been argued to provide crucial organizing structure to linguistic segments in memory (Frazier, Carlson, & Clifton, 2006). A number of studies evidence a direct relationship between prosodic structure and units in memory. For example, the temporal grouping of items leads to better memory for lists generally (Liu & Caplan, 2020). While this benefit is largest for auditorily presented lists with cues to temporal grouping, it is also observed in visually presented lists (Hitch, 1996) that indicate chunks by presenting groups of items together, or by varying the presentation rate of items by including longer pauses at chunk boundaries.

Furthermore, evidence from the list memory literature suggests that prosodic cues to boundaries signal chunks and enhance memory during recall. Frankish (1995) found improved recall for sequences of digits when chunks were introduced via pauses, "natural" intonational groups, and regular stress/pitch changes. Crucially, intonational or melodic contours alone were not sufficient to improve recall accuracy. That is, simply imposing a melody on the list without introducing specific cues to the edges of potential chunks did not improve recall relative to monotone, regularly-timed controls with no variation in inter-item pauses. On the other hand, sequences of monotone syllables with regularly timed stress (e.g. pitch excursions on every fifth syllable) improved recall to the same degree that pauses did, and simultaneously introducing both cues (pauses and stress) did not result in any significant improvement beyond either one on its own. Taken together, these results suggest that prosodic features that serve as cues to boundaries facilitate the chunking process, and when chunks are clearly marked, overall memory for the sequence is improved.

Across a series of experiments, Reeves, Schmauder, and Morris (2000) provide additional evidence that imposing regular stress patterns on lists of words improves recall. The authors also consider the question of whether this is due to increased salience of the stressed items relative to the monotone controls, or to facilitation of perceptual grouping due to the signaling of boundaries, ultimately arguing for the latter. Though Reeves et al. (2000) find improved recall for both stressed and boundary-adjacent items, items in group-final positions tended to show the largest recall advantage, regardless of whether the particular stress pattern was anapest or dactyl. In other words, chunk-internal positions were more prone to errorful recall than chunk-final positions. The authors took this as support for chunk-based models of memory, where interference effects are more likely to obtain within groups than across groups.

A handful of studies have extended the basic grouping effects in word lists to sentence memory. Recall the results of Jarvella's (1971) verbatim recall study discussed in §2.1.1: memory for the medial clause was higher in (19b) than in (19a). This was initially attributed to the fact that in (19b), this clause begins a new sentence.

(19) a. Kofach had been persuaded by the international to stack the meeting for McDonald. The union had even brought in insiders.

b. The confidence of Kofach was not unfounded. To stack the meeting for McDonald, the union had even brought in outsiders.

Notably, the medial clause in the second condition is prosodically separated from the final clause, as is obligatory for subordinate clauses in sentence-initial position. It's then possible that the independent prosodic status of the medial clause in this condition contributed to its higher recall accuracy. To explicitly test the role of prosody, Jarvella (1979) used a similar experimental design, but removed prosodic cues by playing monotone recordings of the sentences read at a controlled pace, thus affecting both the intonational and durational properties of the speech signal. The aim was to remove prosodic cues to clausal/phonological boundaries, and thus eliminate prosodic boundary information as a potential cue to chunking content in memory. Participants still had access to information about clausal boundaries, simply by syntactically parsing the input. The results of these studies showed that the benefit for C2 in (19b) is degraded significantly in the monotone cases, suggesting that prosodic information plays a special role in the structuring and maintenance of content in memory (Marslen-Wilson & Tyler, 1976; Simpson, 2016; Slowiaczek, 1981). Consistent with this interpretation, others (Frazier et al., 2006; Slowiaczek & Clifton, 1980) have argued that finer prosodic structure facilitates more durable memory representations, leading to greater ease of retaining linguistic content in memory. This may partially explain why it is generally easier to remember sentential content compared to disconnected words in a list, though sentential content benefits from multiple sources of hierarchical connectedness (e.g., syntactic, prosodic, event, and discourse structure). In general, memory for prosodically enriched sentences is better than memory for less prosodically articulated sentences, which in turn is better than memory for disconnected word lists.

Simpson (2016) strengthens Jarvella's claim by investigating memory for smaller units of prosodic structure, where prosodic boundaries do not always align with clausal boundaries, using naturalistic examples of phonological phrases[6]. These naturally occurring tokens were taken from the Santa Barbara Corpus of Spoken American English (Du Bois, Chafe, Meyer, Thompson, & Martey, 2000-2005); a sample representative of

[6] *Intonation units* in Simpson's terms, which are intermediate pieces of prosodic structure below the intonational phrase level, but above the word level.

Simpson's stimuli is given in (20). The experiment tested differences in recall accuracy between utterances with low, medium, and high numbers of prosodic units contained within them. Each prosodic unit contained roughly the same number of words, but the total number of words increased across low, medium, and high conditions. For the low and medium conditions, the number of clauses remained roughly the same. Thus, the experiment was designed to test whether short term memory capacity limitations for words increases as the number of prosodic phrases increases. Crucially, many of the syntactic and prosodic boundaries did not align in Simpson's stimuli (see the discussion of the prosodic structure of (20) in §2.1.2.1). Syntax-prosody mismatches like this one were present in Simpson's stimuli at both sub-clausal and clausal boundaries.

(20) (Anyway), (this cat must only weigh like), (a few pounds).

The results suggested that prosodic phrases improve verbatim recall of utterances up to a certain limit (3-6 intonation units), beyond which additional prosodic units do not increase memory capacity any further. Advocating for a stronger prosody-memory relation than earlier studies, Simpson (2016) argues that intonation units but not clauses are the relevant level of representation that structures memory, based on the fact that the results of the model reported did not suggest a significant effect of clauses. The authors suggest that previous effects of clause boundaries on memory may be due to significant overlap between clause boundaries and prosodic boundaries in "written-style language". For auditory stimuli with prosodic boundaries below the clause level, Simpson's claim is likely true; however, that does not rule out the significance of the clause for memory representations altogether in the absence of lower-level explicit cues to prosodic structure.

Neither Jarvella (1971) nor Simpson (2016) test memory for prosodically separated vs. incorporated units that are otherwise lexically and syntactically matched. While there are advantages to using naturally occurring stimuli, a more carefully controlled experiment would also be informative. Ideally, Simpson and Jarvella's findings should be replicable across many different tasks and experimental designs, and using prosodic units of different sizes, in order to support the argument that prosodic structuring generally aids memory. The recognition memory[7] study reported in Experiment 1 (§3.2) partially addresses this.

Studies on implicit prosody have also investigated the relationship between prosodic phrases and working memory capacity. In an individual differences study, Swets et al. (2007) find that readers with low working memory spans are more likely to show a high attachment preference for RC attachment ambiguities, such as *The maid of the princess who scratched herself in public was terribly embarrassed.* This is surprising under an account where low attachment is the less costly syntactic choice (Frazier & Fodor, 1978). The authors suggest that this may be explained by the fact that working memory capacity determines the size of processing units, and that these units may be constructed according to the implicit prosodic structure assigned to a particular sentence. That is, the tendency to package less content within a phonological phrase may increase the probability of positing a prosodic break prior to the relative clause, which in turn may lead to a greater tendency to attach high.

In a similar vein, Kroll and Wagers (2019) put forward a prosodic explanation for the observation that in reading, restrictive relative clauses (RRCs) seem to incur more processing difficulty than appositive relative clauses (ARCs), the focus of Chapter 3 (§3). In line with the proposals put forth by Hirotani et al. (2006) and Swets et al. (2007),

[7] The focus of many early recognition studies on language investigated the loss of "surface" syntactic features compared to the message-level content of sentences (Anderson & Paulson, 1977; Graesser & Mandler, 1975; Sachs, 1967, 1974). In contrast, relatively few studies have investigated prosodic structure using recognition memory (Cohen, Douaire, & Elsabbagh, 2001; Speer, Crowder, & Thomas, 1993). The studies that have done so specifically address the effect of syntax-prosody mismatches on recognition memory for auditorily and visually presented sentences. This research question differs from the current focus, which is to understand the extent to which licit prosodic boundaries cued by punctuation in reading are utilized in the partitioning of linguistic content in memory. Thus, Experiments 1 and 6 support the validity of this methodology for studying effects of implicit prosody.

they suggest that constituents grouped together in the implicit prosodic representation are more likely to share working memory resources. Therefore in prosodically segmented structures like ARCs, working memory resources may be allocated unit-by-unit, as opposed to being distributed across all sentence content as a whole in the RRC case. This particular study, as well as the processing profile of appositives, is discussed in more detail in Chapter 3 (§2.1.2.1). Overall, various sources of evidence support a tight linking between prosodic segmentation and memory units.

2.1.3 Discourse Units

Empirically establishing the relationship between discourse segments and units in memory is difficult, because in most cases, it is not possible to disentangle discourse segments from syntactic or prosodic ones. Nevertheless, discourse segments are formally modeled as important organizing units in language, because relationships between segments have been argued to progress conversations in a systematic way. In line with this view, a handful of proposals discuss the discourse-memory correspondence. This section discusses two approaches to modeling discourse structure in formal linguistics: Question Under Discussion (QuD) theory and Segmented Discourse Representation Theory (SDRT). Each of these frameworks assumes that discourse structure is hierarchical, like other levels of linguistic structure. Both frameworks have also been applied to the processing of appositive relative clauses. Those researchers who posit explicit discourse-memory relationships, however, do not situate their approaches within either of these formal systems. To that end, §2.1.3.2 discusses psychological approaches to modeling discourse structure, which share some notable commonalities with linguistic models. Nevertheless, it is intuitively plausible that linguistically grounded discourse segments can have the same organizational role in memory as proposed in other work on non-linguistically defined segments.

The purpose of the current section, then, is to provide a high-level overview of different approaches to modeling discourse, specifically with respect to (i) how segments relate to each other and (ii) how segments are represented in memory.

2.1.3.1 Linguistic Approaches to Discourse Structure

Researchers propose that discourse segments are propositional units of language that relate to one another in one of two ways: through the Question Under Discussion structure (Roberts, 1996/2012), or through coherence relations between discourse segments (Asher & Lascarides, 2003). I discuss each of these approaches in turn.

The QUD structure of a discourse broadly serves to organize the goals of a conversation, through a means that directly relates to the information structure of utterances. The idea is that conversations are guided by a larger goal: to answer the overarching question, *What is the way things are?* By dividing this larger goal into series of smaller subquestions, conversational participants seek to collaboratively uncover information to be added to the common ground, a set of shared, accepted propositions (Stalnaker, 2002). Through relating each utterance in a conversation to a guiding, implicit question, the model is able to derive the information structural partitioning of utterances, i.e., that information which is in the foreground and is directly associated with the implicit question through *linguistic focus* (see §4.1.1 for more detail), and that information which is in the background. In (21), the focus of the utterance (*the* BEANS) is directly associated with the *wh*-element of the subQUD, *What did Fred eat?*

In this way, a speaker may opt for the strategy of breaking down a complex question like *Who ate what?* into subquestions, like in the contrastive topic structure in (21), or through an alternative strategy (sequentially addressing the questions: *Who ate the rice?, Who ate the beans?, etc.*). In either case, there are distinct prosodic correlates associated with the elements marked as contrastive topics (*Fred* and *Mary*), the portions of each utterance present in the subQUDs, and those marked as foci (*beans* and *rice*), the portions of the utterances associated with the *wh*-element of the subQUDs (see Büring, 2003, for an in-depth discussion).

(21) Contrastive Topic QuD Structure (Büring, 2003)

Who ate what?

What did Fred eat? What did Mary eat?

FRED$_{CT}$ ate the BEANS$_{F}$. MARY$_{CT}$ ate the RICE$_{F}$.

Because this theory is closely tied to the semantics of questions and focus and thus relies on a formally defined notion of relevance, it is restrictive in the discourse moves it allows. In real conversations, participants rarely follow such strict guidelines. Conversations can be rather unstructured, and while there are general principles that govern conversational moves (e.g., Grice's (1975) Cooperative Principle and Maxims of Conversation), it is the job of one's interlocutors to infer how conversational moves relate to the larger goals of any given discourse. As such, more recent extensions of QuD theory have relaxed some of its formal requirements (see Onea, 2016, and Riester, 2019, for example).

SDRT takes a different approach than QuD theory. Asher and Lascarides (2003) (A&L) propose that the central aim of conversation is to understand how units of language cohere with one another. They observe that the relationships between segments of discourse are often inferred in the absence of particular linguistic cues, in accordance with Gricean Maxims. Thus, they posit a defined set of rhetorical relations that hold between propositions (24), which are meant to link the meanings of independent units of discourse together such that they make sense as a whole and constrain the interpretation of certain linguistic elements, like anaphors. The rhetorical relations proposed by A&L are meant to capture how comprehenders jointly recruit their linguistic and world knowledge in order to understand the relationships between segments of language and reconstruct the narrative structure of a discourse. For example, we are able to infer that the sequence in (22) describes a particular temporal relationship between events, one where the segments occur sequentially, although this is not made obvious by available linguistic cues. Though all of these sentences are in the past tense, the order they occur in in the text signifies their order in time. However, ordering of a text does not always map onto order in time. In (23), the second segment is understood as preceding the first in time although it follows it in the text (termed a backshift). (23) also allows comprehenders to infer a causal relationship between segments: that the second event (of pushing) caused the first event (of falling) to occur.

(22) Pierre entered the living room. He sat down on the sofa. He fell asleep.

(23) Max fell. John pushed him.

The set of relations proposed by Lascarides and Asher (1993) is given in (24). Their application to an example discourse is in (25).

(24) Rhetorical Relations (Lascarides & Asher, 1993)

a. Explanation(α,β): the event described in β explains why α's event happened (perhaps by causing it).

b. Elaboration(α,β): β's event is part of α's (perhaps by being in the preparatory phase).

c. Narration(α,β): The event described in β is a consequence of (but not strictly speaking caused by) the event described in α.

d. Background(α,β): The state described in β is the 'backdrop' or circumstances under which the event in α occurred (no causal connections but the event and state temporally overlap).

e. Result(α,β): The event described in α caused the event or state described in β.

(25) $_{\pi 1}$Max had a lovely evening. $_{\pi 2}$He had a great meal. $_{\pi 3}$He ate salmon. $_{\pi 4}$He devoured cheese. $_{\pi 4}$He won a dancing competition.

In analyzing the structure of a discourse, SDRT makes a division between coordinate relations and subordinate ones. Coordinate relations (*Contrast, Parallel, Narration*) are "forward-moving" in that they have the ability to progress a discourse by attaching incoming discourse units to the most recently added segment in the discourse structure, which in turn affects accessibility of preceding discourse units (e.g., for anaphora). Subordinate relations (*Elaboration, Explanation*) lead to hierarchical, embedded discourse structures that do not push the discourse forward. In example (25), π3 and π4 are coordinating units linked via a Narration relation, whereas π1 and π2 are subordinating ones linked via Elaboration. This division is particularly relevant to the analysis of appositives, which are often argued to contribute discourse-subordinate content, a property that has been leveraged in the processing literature; I return to a discussion of this in §3.1.2.2.

In sum, different theories differ in (i) their formal assumptions, (ii) their means of relating discourse units to one another, and (iii) the empirical phenomena they seek to model. Some researchers propose hybrid theories, which involve a correspondence between SDRT relations and QuD structures (Jasinskaja, 2016; Riester, 2019): that within an SDRT framework, each discourse unit raises an issue, which corresponds with a QuD. In this way, both guiding questions and coherence relations may jointly contribute to the progression of a discourse.

Each of these theories has been used to describe the discourse status of appositive relative clauses, which are argued to contribute secondary information relative to main clause content, either through irrelevance to the main QuD or through a discourse-subordinate relation, in the sense of A&L (1993). I return to a detailed discussion of this in Chapter 3 (§3). For current purposes, it is sufficient to note that a large body of theoretical work posits that the segmentation of roughly proposition-sized linguistic units is relevant to the comprehension of discourse. Neither QuD Theory nor SDRT explicitly posits a relationship between discourse segments and units in memory, but they very naturally lend themselves to this assumption and closely resemble work on the discourse-memory relationship in the psychology literature. The proceeding section discusses psychological approaches to modeling discourse.

2.1.3.2 Psychological Approaches to Discourse Structure

One influential psychological model of the discourse-memory relationship is Situation Model Theory (SMT; Zwaan & Radvansky, 1998). Like SDRT, SMT aims to understand how segments of a discourse are integrated, but also explicitly considers how the structure of a discourse in memory affects storage and retrieval of content from previous segments. Broadly, SMT assumes that syntactic and semantic aspects of language are sets of "processing instructions" for how to construct mental representations of situations. The model posits five situational dimensions relevant to discourse processing: temporal, spatial, causal, motivational, and entity-related goals. Under this model, the current situation model (i.e., the one actively being constructed based on the content of the current clause) occupies the contents of working memory, whereas the integrated situation model occupies "long-term working memory", a separate store within long-term memory that contains relevant portions of previously processed text and is relatively more accessible than the rest of the contents in long-term memory. The contents of short-term working memory contain retrieval cues to the integrated model, which are used in *updating*, or the process of forming links between the current situation model and the integrated situation model. In other words, a situation model is constructed based on the event described by the current clause, and then links are formed between the current and integrated models along each of the situational dimensions (Zwaan, Langston, & Graesser, 1995). Once the current model has been integrated and the integrated model has been updated, the construction of a new current situation model begins.

The process of transitioning to a new current situation model is thought to be costly in terms of the cognitive resources it requires. In cases of significant discontinuities in the situational dimensions that comprehenders monitor during discourse processing, updating of the situation model creates a boundary in the discourse representation. For example, sequences of sentences describing continuous spatial relations (*the knife is in front of the pot, the pot is behind the dish, the dish is to the left of the glass...*) are read faster than those that describe discontinuous spatial relations (*the knife is in front of the pot, the glass is behind the dish, the pot is to the left of the glass*) (Ehrlich & Johnson-Laird, 1982). The authors interpret facilitation of reading times in the first case as a consequence of spatial proximity; that is, content in the current model that is perceived as more continuous with content in the integrated model should be easier to integrate[8]. Similarly, in experiments investigating the processing of time shifts, discontinuities on the temporal dimension evidence boundaries in the mental representation of situations and their organization in memory. In cases where time shifts were too long to be considered part of the initial situation (e.g., 6 hours later in a movie-watching situation), time to identify referents from the initial situation was longer than with time shifts within the confines of the initial situation (10 minutes later). Similarly, probe recognition response times for previous words following a short shift are faster than that of a long shift (Zwaan, 1996). The authors take these facts as evidence that spatial and temporal discontinuities lead to boundaries in

[8]Note that this just as easily could have been because the organization of clauses in the continuous condition allowed participants to form stronger associations between the clauses themselves (see §2.3.2.1).

discourse representation that lead to the construction of a new situation model.

SMT implicitly assumes that discourse segmentation occurs in a clause-by-clause manner, but underspecifies the status of previous discourse segments (those that are not the current situation model) in long-term memory. Some studies propose only a distinction between the current model and the integrated model, which does not preserve the boundaries of previous segments. Following Chafe's (1979, p. 180) suggestion below, the authors assume that the situational dimensions of text comprehension override the presence of relatively more ephemeral segment boundaries.

> Rather than think of an experience as being stored in memory in terms of distinct episodes, it seems preferable to think of a more complex storage in terms of coherent spaces, coherent configurations of characters, coherent event sequences, and coherent worlds. At points where all of these change in a maximal way, an episode boundary is strongly present. But often one or another will change considerably while others will change less radically, and all kinds of varied interactions between these several factors are possible.

Chafe's point is well taken; there are indeed various dimensions along which comprehenders may choose to organize information in memory, as discussed at the outset of this section. Overall, though, empirical evidence used to bolster SMT suggests that discontinuities on spatial and temporal dimensions trigger boundaries in the mental representation of a discourse, which has consequences for the accessibility of content in previous segments. Therefore, SMT proposes a correspondence between discourse segments and the organization of memory. It is conceivable that the frameworks discussed in the previous section (QuD theory and SDRT) could guide a similar relationship between segments of linguistic structure and units in memory. Two other accounts are evocative of this idea as well and make specific claims about the accessibility of parenthetical content in memory. These proposals are reminiscent of recent psycholinguistic accounts of the processing of appositives (see §3.1.2) and offer a useful starting point for the mechanisms considered in Chapter 3.

Grosz and Sidner (1986) propose that discourse processing depends on three components: (i) a linguistic structure consisting of the utterances of a discourse, which naturally organize into discourse segments, (ii) an intentional structure, representing the broad goal (discourse purpose) and subgoals (discourse segment purposes associated with each segment) of a discourse, and (iii) a dynamic attentional structure, akin to the focus of attention, which tracks information about entities, properties, and relations that are most salient at any given point. The intentional structure is assumed to be hierarchical in nature, as is assumed for theories of discourse in formal linguistics. Grosz and Sidner propose that the cues used to diagnose linguistic boundaries may be lexical (*incidentally…*), syntactic (*in the first place…*), or prosodic in nature (the duration of between-segment pauses (Chafe, 1980) and speech rate increase from the beginning to the end of a segment (Butterworth, 1975)). This underscores the fact that the prosodic structure and discourse structure are inherently linked.

Grosz and Sidner's attentional structure contains a set of focus spaces called the focusing structure. Each focus space corresponds to a particular discourse segment and contains the salient entities of that segment. The focusing structure is modeled as a push-down stack ordered relative to the hierarchical relationships between discourse purposes in the intentional structure. Thus, content contained within lower focus spaces is accessible but less salient than content in higher ones. In their analysis of interruptions, parenthetical segments like the italicized content in (26), G&S suppose that there is a "return" to the interrupted discourse segment. They model this in the following way: by popping the focus space associated with the interruption off the stack once it is complete, the next available focus space at the top of the stack will be D1, the pre-interruption segment, because its discourse segment purpose has not yet been satisfied.

(26) $D1_a$: John came by and left the groceries

D2: *stop that you kids*

$D1_b$: and I put them away after he left

This example highlights that G&S's framework is not a sufficient incremental theory of discourse processing. From an incremental perspective, there is no immediate linguistic cue that the incoming segment after the interruption will be a continuation of D1 rather than the start of a new segment. Consider, for example, the alternative continuation: *And I took the dog on a walk.* On a view where segments roughly correspond to clauses in the syntax, and the event described by this clause sequentially follows the one in D1, this should correspond to a distinct segment D3 rather than a continuation of D1. In (26), the linguistic cues indicating a relationship between $D1_a$ and $D1_b$ are the anaphors *them* (= the groceries) and *he* (= John), which make reference to entities in the initial segment but do not occur until later on in the clause corresponding to $D1_b$. Thus, it's not clear from the outset of the post-interruption content whether the speaker intends to start a new discourse segment or to add to the contents of D1. Abstracting away from particular linguistic details, this is exactly the shape of the processing puzzle posed by appositive relative clauses: mechanistically, how is it that pre-appositive content is reinstated such that it can be added to? G&S don't address this particular issue, but they do provide a set of lexical and syntactic cues to particular focus space changes and hierarchical relationships (p. 198).

Redeker (2006) similarly focuses on the role of attentional cues to discourse segment transitions and more explicitly discusses the issue of when a comprehender may choose to return to a previous discourse segment instead of beginning a new one. Under this approach, prosodic cues (the intonational contour, as well as shifts in speech rate, pitch, or volume) mark discourse segment boundaries. Like Grosz and Sidner, Redeker assumes that encountering such a boundary closes the current focus space and opens a new one in the default cause. Then, a set of lexical cues, termed *discourse* *operators* (27), allow a comprehender to identify particular rhetorical relationships between discourse segments.

(27) **Discourse operator**: any expression that is used with the primary function of bringing to the listener's attention a particular kind of relation between the discourse unit it introduces and the immediate discourse context (Redeker, 2006).

In the context of particular discourse operators marking a parenthetical unit, a comprehender may choose to suspend the current focus space instead of closing it off. Redeker (2006) hypothesizes that if a focus space is closed, referents associated with that space should be deactivated. On the other hand, if a space is suspended, its referents should be rendered relatively less active but still accessible. Then, if the segment following a parenthetical is a continuation of the initial segment, a "pop-up" discourse marker (like *but anyway...*) may cue this, and lead to reactivation of the initial segment. Of course, a return to a previous segment need not always be lexically cued by a pop-up marker. In this case, the comprehender should prefer to open a new focus space for the incoming segment.

Evidence for this account came from a cross-modal naming experiment, where listeners heard passages of naturally-occurring speech with discourse markers inserted or removed at segment boundaries. At various points throughout each passage, participants were asked to name a visually presented word aloud, and their naming latency was measured. These words were semantically related to the contents of previous segments that varied in discourse status (parenthetical vs. non-parenthetical) and the presence of a discourse operator (present vs. absent). Redeker hypothesized that naming latencies should be shorter for content semantically related to a suspended prior segment compared to a closed prior segment. The results suggested that the presence of an overt operator for non-parenthetical preceding segments had no effect on naming latency. For preceding parenthetical segments, however, the presence of push-markers (overtly cueing the beginning of a parenthetical) and pop-markers (cueing the end of a parenthetical) resulted in facilitation of naming latency relative to identical segments where those overt markers were missing. The results were taken as support for the claims that (i) the default preference is to begin a new segment, (ii) overt discourse cues to suspend a segment (rather than close it) facilitate access to semantic content from that segment, and (iii) overt discourse cues to reinstate a suspended segment facilitate access to semantic content from that segment as well.

Redeker further surmises that the effect of a particular marker is a function of its discourse context and not its lexical semantics, as some of the operators marking parenthetical and non-parenthetical transitions were lexically identical. Again, this raises the question of when during the processing of a discourse particular focus space operations (suspension and reinstantiation) occur, and how much discourse context is needed to override the default preference to open a new segment, even in the presence of overt markers. Naturally, the need to wait for further content in order to determine the discourse status of any given segment gives rise to the possibility that segment status (active, suspended, closed) must sometimes be revised at some point after encountering a segment boundary and positing the opening of a new segment. This issue is underexplored in the approaches discussed in the current section, although Redeker's experimental results suggest that non-marked transitions result in longer naming latencies. This affirms the suggestion that the default preference is to close the previous segment and move onto the next one. Like Grosz and Sidner's account, then, Redeker's leaves open some important issues pertaining to the real-time segmentation of linguistic units, as well as when and how the discourse status of particular segments is encoded. I return to a more thorough discussion of this issue in Chapter 3 in the context of appositive relative clauses.

2.2 Traversing Boundaries Online

Up until this point, I've summarized evidence that segmentation in linguistic memory occurs across multiple levels of analysis. Two remaining questions concern (i) how those segments are formed in memory (the focus of §2.3.2) and (ii) whether such segmentation affects accessibility of linguistic content during comprehension (§2.2.2). Very little work in sentence processing explicitly contends with the second question. In the current section, I review existing work that does address this question, which is situated in two distinct but overlapping areas of investigation.

The first area concerns how the presence of prosodic boundaries affects the organization of prosodic phrases in memory, and how parsing decisions and memory retrieval operations proceed over those boundaries. For online dependencies spanning prosodic boundaries, some work on Prosodic Visibility effects suggests that boundaries have a detrimental effect on accessibility of content in previous prosodic phrases. Other studies fail to find such visibility effects. The second area concerns the unique processing profile of appositive relative clauses, constructions in which the presence of sentence-medial prosodic (and discourse) boundaries seems to facilitate access to linguistic content in certain segments, an effect termed *discounting* (Duff et al., 2023). Chapter 3 will review a series of experiments that fail to attribute appositive discounting to discourse-based sources, leaving syntactic or prosodic independence as potential contributing factors to discounting effects.

At face value, it would seem that the effects of ARC boundaries and other intermediate prosodic boundaries are at odds with one another. This apparent conflict warrants more careful consideration of previous work on the role of prosodic boundaries. To that end, this section introduces the Prosodic Visibility Hypothesis (§2.2.1), and then reviews evidence previously leveraged to support and refute it (§2.2.2). This discussion highlights a secondary issue that this section aims to address: the assumptions about segment-level memory adopted in the sentence processing literature are often imprecise and inconsistent. This is in part because there is an active debate in sentence memory literature about the maintenance of linguistic content in memory. It then compares mixed results from the Visibility literature with the robust and reliable ARC discounting effect observed elsewhere.

2.2.1 The Prosodic Visibility Hypothesis

§2.1 established the case that greater prosodic segmentation within a sentence (i.e., the presence of sentence-internal prosodic boundaries) has a beneficial effect on offline memory (Jarvella, 1979; Simpson, 2016, i.a.). In contrast, some studies on prosodic parsing report that boundaries have a detrimental effect on access to pre-boundary content during comprehension. This has led to conflicting claims about the relationship between prosodic segmentation and resulting memory representations in the sentence processing literature. Some studies propose that previously built syntactic structure and syntactic features in memory become decreasingly accessible to the parser as a function of the number of intervening prosodic boundaries from the phrase currently being processed. This has been termed the prosodic Visibility Hypothesis (VH). The original framing of VH, proposed by Schafer (1997), is stated in (28).

(28) **Prosodic Visibility Hypothesis** (VH) (Schafer, 1997)

a. The phonological phrasing of an utterance determines the *visibility* of syntactic nodes.

b. Nodes within the phonological phrase currently being processed are more visible than nodes outside of that phonological phrase; visibility is gradient across multiple phonological phrases.

c. In first analysis and reanalysis, attachment to a node with high visibility is less costly in terms of processing/attentional resources than attachment to a node with low visibility.

Schafer (1997) defines visibility in terms of which syntactic nodes are least costly for the parser to access during incremental syntactic attachment decisions. This conception of the VH assumes that phonological phrases structure sentences in memory such that accessing content in the phrase currently being built is relatively low-cost, whereas accessing content in previously completed phrases becomes increasingly more costly as the distance between the current phrase and the target phrase increases.

A schematic representation is given in (29), where the value associated with each prosodic phrase represents the visibility of that phrase, in terms of hypothetical levels of activation, at the time of processing the final segment (*with a mean look*) which may be attached to the verb (*angered*) or NP2 (*the rider*). The VH assumes that attachment of this PP to NP2 (*the rider*) will be easier than attachment to the verb (*angered*), because NP2 is relatively more visible. Of course, items are assumed to have varying activation levels of their own, but Visibility supposes that all else equal, content two phrases back from the current timestep will be less accessible to the parser than content one phrase back. Thus, this framing of Visibility is distinct from a pure recency-based account of accessibility, because if two candidates for retrieval or attachment are contained within the same prosodic phrase, they should be equally "visible", and perhaps phrase-level visibility gates the the effect of item-level variance in activation (see Van Handel, 2022).

(29) Hypothetical Visibility by Phrase

$_\varphi$ the bus driver angered	$_\varphi$ the rider	$_\varphi$ with a mean look
0.4	0.6	0.8

There are multiple means of spelling out the mechanism underlying visibility-like effects. At a minimum, the VH assumes direct access to content within the current prosodic phrase; however, Schafer's original definition of Visibility does not explicitly comment on the status of previous phrases in memory.

In subsequent work on Visibility, Frazier and Clifton (1998) and Carlson et al. (2009) suggest a loss-based view of Visibility: that the strength of the representation of each prosodic phrase in memory progressively decreases as a function of its distance from the currently being processed prosodic phrase. Such a view implies that processes at prosodic boundary positions result in loss of previous content in memory. Put differently, this view assumes that boundaries trigger degradation of previous content, through (i) active suppression of syntactic structure in memory, or (ii) decay of syntactic detail due to the formation of more durable segments at higher-order conceptual levels of representation (Potter & Lombardi, 1990).

It is notable that this particular implementation of the VH is in conflict with offline studies on prosodic memory. If encountering a prosodic boundary leads to degradation of syntactic memory during incremental comprehension (relative to non-segmented sentences), it would then be mysterious that offline sentence memory for prosodically segmented sentences ends up being better than that of non-segmented sentences.

Schafer's own definition of Visibility does not necessarily entail loss. The VH simply states that attaching to a more visible node consumes fewer "processing or attentional resources". One version of this hypothesis then allows for the idea that memory for previous content is not compromised by crossing a prosodic boundary during the incremental parse. It may instead be that prosodic phrases group syntactic content in memory, and that these groups are subject to the usual constraints on memory in that only the currently being constructed phrase is being actively maintained, whereas previous segments are accessible only if their contents are reactivated. Insofar as the parser prefers following the path of least resistance, it then defaults to attaching within the current phrase, when possible, so as to avoid performing unnecessary retrieval operations. Under such a view, when there is no licit attachment site within the current prosodic phrase, the parser should be more likely to attempt to reinstate a previous phrase in order to search its contents for an appropriate attachment site. Generally speaking, this particular version should predict only a division in accessibility between the currently being processed phrase and all earlier phrases, without adopting a gradient notion of accessibility[9]. This is essentially a recency-based view of Visibility, but one that determines recency on the basis of phrase-membership.

Crucially, the VH as stated in (28) is compatible with either of these approaches (i.e., loss-based or recency-based Visibility). The following section reviews evidence for Visibility, establishing that previous work on attachment decisions provides little evidence to support a loss-based view. While there is little dispute surrounding the claim that the position of an earlier prosodic boundary mediates syntactic attachment, studies that investigate the interaction between attachment preferences in the context of multiple preceding boundaries do not yield conclusive results. Therefore, a recency-based view of Visibility is better-supported. §2.3.3 will return to how this might relate to the organization of prosodic phrases in memory.

2.2.2 A Murky Empirical Landscape

Schafer's (1997) evidence for Visibility comes from patterns of offline interpretations for attachment ambiguities. Using auditory stimuli like in Table 2.1, Schafer (1997) finds that increasing the number of intervening boundaries between a PP (*with a mean look*) and two possible attachment sites influences the mostly likely interpretation of the sentence. Phonological phrase boundaries are indicated by the ($_\varphi$) groupings in

[9]It's possible that in the absence of a cue to a particular segment, previous segments may be reinstated sequentially from most to least recent. However, the weaker backwards contiguity effect in list and sentence memory suggests that backwards serial search of phrases in this manner is not likely.

Boundaries	Sentence	% VP
-NP2, -PP	($_{\varphi}$The bus driver angered the rider with a mean look)	60%
+NP2, -PP	($_{\varphi}$The bus driver angered) ($_{\varphi}$the rider with a mean look)	44%
-NP2, +PP	($_{\varphi}$The bus driver angered the rider) ($_{\varphi}$with a mean look)	62%
+NP2, +PP	($_{\varphi}$The bus driver) ($_{\varphi}$angered) ($_{\varphi}$the rider) ($_{\varphi}$with a mean look)	52%

Table 2.1: Example item set from Schafer (1997), Experiment 1. +/- Boundaries indicates the presence vs. absence of pre-NP2 and pre-PP boundaries. The %VP column indicates the rate of VP-attachment preference by condition.

Table 2.1, and offline attachment preferences are reported in the %VP column.

The -NP2,-PP condition exemplifies a standard PP-attachment ambiguity configuration, where the prepositional phrase *with a mean look* can attach to either the VP (*angered*) or NP2 (*the rider*). When attached to the VP, the resulting interpretation is one in which the bus driver has a mean look which causes the rider to be angry. When attached to NP2, the rider is the one who has a mean look. Schafer's design varied the presence vs. absence of a pre-NP2 boundary and a pre-PP boundary. In addition, the +NP2, +PP condition contained an additional post-NP1 boundary in order to avoid any unnaturalness from unevenly sized prosodic phrases, as (*the bus driver angered*), if phrased together, would have been noticeably longer than (*the rider*).

The auditory stimuli were produced with H* pitch accents within each phonological phrase, and L- phrasal tones at the right edge of each phrase; ToBI transcriptions of these stimuli are provided in (30). Phonological phrase boundaries were manipulated via the presence of L- phrasal tones. While Schafer does not explicitly comment on the presence of pauses at these phrasal boundaries, visual inspection of the waveforms presented in Schafer's Figure 2.1 suggest that phrasal tones in these stimuli did not always align with prosodic breaks. Thus, it is unclear the extent to which the results are driven by intonational cues versus durational cues.

(30) a. (The bus driver angered the rider with a mean look)
H* H* H* H* H* L-L%

b. (The bus driver angered) (the rider with a mean look)
H* H* L- H* H* H* L-L%

c. (The bus driver angered the rider) (with a mean look)
H* H* H* L- H* H* L-L%

d. (The bus driver) (angered) (the rider) (with a mean look)
H* L- H* L- H* L- H* H* L-L%

Contra many other studies that find a default low attachment preference in English, Schafer (1997) found a default VP-attachment preference in the -NP2, -PP condition, where there were no sentence-internal phonological phrase boundaries. The preferred interpretation for the -NP2, +PP condition, which contained a pre-PP boundary, tracked the baseline condition. This is consistent with much other work suggesting that a boundary following a potential attachment site serves to repel attachment to that site. It is important to note that Visibility alone cannot explain this preference; however, if we further suppose that prosodic boundaries are a proxy for syntactic grouping (Harrington Stack & Watson, 2023; Van Handel, 2022), a pre-PP boundary in this condition effectively blocks NP2 attachment. In the +NP2, -PP condition, Schafer found a low attachment preference. This is consistent with the VH, as NP2 (*the rider*) is contained within the same phonological phrase as the PP. Finally, the +NP2, +PP condition provided the critical evidence for Visibility, which predicts that an attachment site two prosodic groups back should be less visible than a site one prosodic group back. This condition revealed that attaching to the VP is relatively more dispreferred given two intervening φ-boundaries compared to +NP2, -PP, where there is only one intervening φ-boundary. Thus, the results were taken to provide support for the Prosodic Visibility Hypothesis, because high attachment was dispreferred when the VP attachment site was packaged in a separate prosodic phrase and there was a more accessible attachment site available to the parser.

Other studies have failed to extend the basic visibility effect reported by Schafer (1997). Carlson et al. (2009) sought to determine whether visibility extends to depen-

Correlate	1 ιP-Bound	2 ιP-Bounds
Main	($_\iota$Diane thought that the movie was well-edited) ($_\iota$not Louisa)	($_\iota$Diane thought) ($_\iota$that the movie was well-edited) ($_\iota$not Louisa)
Embed.	($_\iota$Diane thought that the movie was well-edited) ($_\iota$not the soundtrack)	($_\iota$Diane thought) ($_\iota$that the movie was well-edited) ($_\iota$not the soundtrack)

Table 2.2: Example item set from Carlson et al. (2009), Experiment 2b.

dency resolution more generally, not just syntactic attachment. In a listening study, Carlson et al. (2009) investigated the resolution of replacive ellipsis (*Diane likes sci-fi, not* <~~*Diane likes*~~> *fantasy*) across intervening intonational phrase boundaries. This type of ellipsis involves establishing a contrastive relationship between a remnant (*fantasy*) and correlate (*sci-fi*), then eliding non-contrasting material.

The design crossed the position of the Antecedent (Matrix Clause, Embedded Clause) with the number of intervening ι-boundaries (1, 2). An example item set is given in Table 2.2, where intonational phrase boundaries are indicated with parentheses. Recordings of stimuli were produced with varying pitch accents on each of the nouns, and with L-H% continuation rises, pre-boundary lengthening, and pauses at each intonational phrase boundary (see pitch tracks in Carlson et al. (2009) for more detail).

As discussed earlier, Carlson et al. (2009) adopt a version of Schafer's hypothesis that assumes gradient representation of each previous prosodic phrase in memory; that is, memory for previous ιPs is assumed to be degraded relative to more recent ones. Given the hypothesis that intonational phrase boundaries serve to segment sentences in memory, they expected to find an interaction such that matrix clause-correlates with two intervening ιP boundaries should be more difficult to access than those with only one intervening ι-boundary. However, they found only a main effect of antecedent position such that embedded clause correlates were preferred over matrix correlates, but no effect of number of boundaries and no interaction. Similarly, in experiments measuring sentence-final reaction times in a "got it" task, they found no evidence that RTs were modulated by boundary presence.

Carlson et al. (2009) adopted a semantic analysis of replacive ellipsis constructions following Reinhart (1991), where ambiguous cases (*Diane thought that Felicia was well-dressed, not Louisa*) are not due to ambiguities in phrase structure. Instead, they assumed that the replacive structure involves root adjunction and that the choice of correlate is determined at Logical Form. Carlson et al. ultimately suggested that the information structure of preceding material was more important for resolving replacive ellipsis than the position of boundaries, as replacives are a type of focus-sensitive ellipsis, in which the correlate is typically prosodically marked with an accent. Thus, they assume that determining the correlate-remnant pair is not a syntactic process but is thought to depend on the semantic and information structural properties of the sentence. In line with this analysis, follow-up experiments suggested that accent position plays a larger role in the search for a correlate in these constructions than boundary presence and position. That being said, it is somewhat misleading to suggest that replacive ellipsis does not involve a syntactic dependency at all; after all, the search for a correlate is guided by syntactic properties of the remnant, and retrieval of syntactic content must underlie the reconstruction of replacive ellipsis (e.g., <*thought that the movie was well-edited*> in the Main conditions). Therefore, I assume that resolving replacive ellipsis, in part, requires accessing syntactic memory. I discuss these processes at length in §4.1.1, as Experiments 4-8 utilize similar focus-sensitive ellipsis constructions.

Taking Schafer's results into consideration, they suggest that the role of Visibility is limited to influencing syntactic attachment decisions, but not dependency resolution in general. More generally, they propose a specialized role for prosodic boundaries, such that they do not influence all linguistic dependencies equally. This reasoning is not without consequences for the mechanism driving Visibility. An attachment-resolution asymmetry would be unexpected if a gradient memory-based explanation is correct, as degraded memory for previous prosodic phrases should presumably affect all memory retrieval operations to some extent. Carlson et al.'s proposal also departs from Schafer's original position, which assumes that phonological phrases guide syntactic processing through a visibility-type mechanism, but that intonational phrases constrain semantic/pragmatic evaluation. Assuming such a syntax-semantics division might predict that Carlson et al.'s stimuli would foster the ideal conditions to observe a boundary effect. Nevertheless, their results fail to validate Visibility.

Other work has similarly noted the fact that visibility-based hypotheses often underspecify their mechanistic assumptions or vary significantly from one another. Following up on Frazier and Clifton's (1998) suggestion that a number of factors may influence the accessibility of dependents, Harrington Stack and Watson (2023) attempt to connect segment-based accessibility to perceptual groupings. Specifically, they posit that if prosodic boundaries lead to the formation of "perceptual units", that perceptual breaks should more generally evidence visibility. Like Schafer (1997), they investigate the role of boundaries in attachment ambiguities, as in (31), where *after John visited* could attach to *telephoned* or to *learned*. The '|' symbols in (31) indicate prosodic boundary positions manipulated in their experiments.

(31) Susie learned that Bill | telephoned | after John visited.

Here, they fail to find evidence to support a general perceptual role for prosodic visibility. They first replicated the well-established finding that prosodic boundaries influence interpretation (i.e., that an early boundary influences low attachment, whereas a late boundary influences high attachment). Across two experiments, they then introduced two non-linguistic cues to perceptual groupings: a buzzer sound and an artificially-generated pause in the absence of other prosodic cues to group boundaries, like phrase-final lengthening. They found that these non-linguistic cues to perceptual breaks do not yield prosodic visibility effects. They argue that the role of prosodic boundaries in syntactic parsing is not due to perceptual groupings that form processing units, but instead that boundaries serve as probabilistic cues to different syntactic structures. This explains why a boundary following a linearly closer dependent leads to high attachment, a preference that visibility alone cannot account for.

These results provide evidence against a very particular conception of Visibility, but it may be important to draw a distinction between cues to perceptual breaks and linguistically-defined units in memory. The meaning of the term "perceptual unit" is not well-defined. It would be unsurprising if the presence of noise in a linguistic stimulus is not sufficient to create a unit in memory, especially because listeners likely make use of a combination of prosodic cues in order to gather enough evidence to posit a prosodic boundary. It's also likely that in a sentential context where top-down information about linguistic structure is available, the parser would not take into account a non-linguistic noise or perceptual break as a meaningful cue to group syntactic content in memory. After all, language processing in naturalistic contexts often involves filtering out irrelevant stimuli. In these cases, it may be that structural cues to linguistic units outweigh non-linguistic ones.

Other studies have advocated for different interpretations of Prosodic Visibility, focusing instead on the tradeoff between two factors that mediate accessibility: (i) prosodic grouping of syntactic content and (ii) prominence of accented constituents under linguistic focus. A series of studies on length effects in implicit prosody lend support for the Visibility First Hypothesis (VFH; Van Handel, 2022), as in (32). §4.1.1 discusses linguistic focus in greater depth; for the time being, I explain the predictions of this hypothesis in terms of which noun phrase receives greater prominence (in terms of accenting, focus, modification, etc.) in a general sense.

(32) **Visibility First Hypothesis** (Van Handel, 2022)

The parser preferentially attaches incoming material to the most visible potential attachment site, because prosodic visibility serves as a proxy for syntactic grouping. The parser weighs additional evidence, such as focus, only when two attachment sites are equally visible, because in this case visibility is insufficient for determining the most likely syntactic grouping.

Length effects in RC attachment ambiguities provide critical evidence for the VFH. In (33), the relative clause *who chided herself* is globally ambiguous: it may attach to either N1 (*the niece*) or N2 (*the waitress*). Previous work establishes that modified noun phrases (*the incredibly diligent niece*) attract attachment, perhaps because nouns modified by gradable adjectives receive a contrastive interpretation (Sedivy, Tanenhaus, Chambers, & Carlson, 1999) and are thus more likely to receive an accent in the implicit prosodic representation, or because modification contextually enriches the representation of a noun, thus increasing its level of activation in memory (Hofmeister, 2011). Thus, a modification-based explanation alone would predict a greater preference for N1 attachment for the example in (33b). In contrast, Van Handel (2022) found a greater preference for N2 attachment. If a longer N1 encourages a prosodic break following *niece*, as in (34), this would cause *waitress* to be more visible at the point of parsing the RC, leading to greater accessibility of N2 over N1, despite the prominence of N1. This is exactly what the VFH predicts: that the most accessible syntactic node as determined by prosodic grouping is more likely to attract attachment of incoming constituents.

(33) a. The niece of the waitress [who chided herself]...

b. The incredibly diligent niece of the waitress...

(34) ($_{\varphi}$The incredibly diligent niece) ($_{\varphi}$of the waitress who chided herself)...

Van Handel's (2022) particular approach admits a role for prosodic boundaries in mediating syntactic attachment decisions, but remains neutral on whether such effects are driven by the structure of prosodic phrases in memory. I return to this point shortly.

Finally, I return to the observation about appositives that I began with at the outset of the dissertation: not all sentences with intermediate prosodic boundaries yield Visibility effects. Crucially, the processing profile of ARCs potentially leads to the opposite conclusion, that the presence of preceding prosodic boundaries facilitates parsing and/or dependency resolution (Dillon et al., 2017, 2018; S. Kim & Xiang, 2023). For example, main clause dependencies that span a sentence-medial, prosodically isolated ARC, like the anaphora in (35), are facilitated relative to the same dependencies spanning RRC structures, which lack intermediate boundaries.

(35) **The cat**, who loves to hunt bugs, was chasing **her tail**.

As discussed previously, the investigation of ARCs has proven challenging because they are set apart at multiple levels of linguistic representation. Despite this, a growing body of work has failed to find support for discourse-based explanations of their processing profile (Duff et al., 2023; Kroll & Wagers, 2019). That is, multiple definitions of discourse-independence fail to capture their ability to be bypassed. Although some authors note that their prosodic boundaries may be relevant, only one study (Kroll & Wagers, 2019) has sought to explicitly determine how boundaries relate to their independence. The current empirical landscape should compel researchers to take seriously the other aspects of independence that appositives display in the search for an explanation for this puzzling effect. I will ultimately argue that while their prosodic independence may not solely explain the processing behavior of ARCs, the prosodic boundaries that surround ARCs (and other sentence-medial prosodic phrases) play a crucial role in their ability to be encoded as independent units in memory.

This section has established that linguistic boundaries have an influence on incremental sentence processing. However, there is disagreement in the literature on (i) exactly what influence different types of boundaries have and (ii) how this interacts with linguistic memory. In particular, the studies reviewed here on Visibility – the tendency for prosodic boundaries to hinder cross-boundary dependencies – are largely inconsistent in their claims about the prosody-memory interface. However, two consistent conclusions emerge: (i) that Visibility effects obtain most clearly for syntactic attachment decisions, and (ii) that attachment sites in the current prosodic phrase are most visible. On the other hand, there exists evidence that the presence of sentence-internal prosodic or discourse boundaries leads to facilitation of access to cross-boundary content in one construction, namely ARCs.

At present, it's unclear whether Visibility is due to a property of the organization of prosodic phrases in memory, or to a parsing preference independent of memory. The fact that a late boundary tends to drive high attachment (e.g., that the RC is more likely to attach to N1 in *The daughter of the actress, who embarrassed herself...*) potentially suggests that a learned syntactic parsing strategy to avoid boundary-adjacent sites underlies some Visibility-related effects. Alternatively, the Visibility First Hypothesis (Van Handel, 2022) suggests that N1 and N2 would be equally visible in this case, and that other factors modulating accessibility determine the ultimate attachment preference here. This would be consistent with a memory-based explanation. Additionally, it remains unclear whether Visibility applies to item-to-item dependencies (like agreement, ellipsis, anaphora, etc.); a memory-based view predicts that it should.

Overall, the landscape of Visibility effects is both empirically and theoretically murky. Something more must be said about why it is that some prosodic boundaries facilitate access to pre-boundary content, whereas other boundaries hinder access.

One aim of the dissertation is to clarify this asymmetry. The following section briefly comments on how a memory-based account could capture Visibility, but Experiment 4 (§4.1.1) and Experiments 7-8 (§5.6) attempt to re-examine this issue in more detail. Ultimately, the results point to an interaction between the position and number of earlier prosodic boundaries, and the particular dependency at hand.

2.3 Memory Mechanisms

Ample evidence across the sentence memory literature supports the position that linguistic units correspond to units in memory. Despite this, the role of segmentation is underexplored in psycholinguistic literature on memory retrieval of linguistic items during incremental comprehension. This section aims to unite these two literatures. I first review the *cue-based retrieval* (CBR) model often adopted in sentence processing in §2.3.1, and then turn to a review of the Temporal Context Model (TCM; Howard & Kahana, 2002) in §2.3.2. §2.3.3 then introduces the Context-Sensitive Encoding Hypothesis, which lays out a proposal for how linguistic boundaries may affect both off- and online sentence memory using elements of CBR and TCM.

2.3.1 Preliminaries

Following a substantial body of work in psycholinguistics, I assume that *cue-based retrieval* – a direct access, content-addressable memory retrieval mechanism – underlies the formation of item-to-item dependencies during incremental sentence comprehension. It is generally assumed that as the words of a sentence come in and out of attention, each item is encoded in memory and incorporated into the current syntactic parse of the sentence. At certain points during processing, successful completion of a parse and computation of meaning requires accessing previously encountered content from memory; it is at these points that retrieval of items that were encoded earlier in a sentence or discourse must be retrieved from long-term memory and reactivated in order for further processing operations to transpire successfully. Studies on cue-based retrieval in sentence contexts have found that just like domain-general retrieval processes, linguistic retrieval is subject to *interference effects* in the presence of cue overload (Anderson & Bower, 1974). That is, when there are multiple feature-matching items in memory, retrieval of a target item becomes more difficult, because activation is spread across feature-matching items thus lowering the activation of any given retrieval candidate. Because the retrieval mechanism searches for feature-matching content in memory in a content-addressable, direct-access manner, it is sometimes error-prone. Therefore, some proportion of the time, retrieval of a partially feature-matching but otherwise incorrect item occurs.

Evidence for the existence of similarity-based interference effects has been gathered across a variety of linguistic dependencies and types of linguistic features. In one investigation, Van Dyke and Lewis (2003) found that the presence of partially matching semantic cues interferes with the process of subject-verb agreement and integration. Their stimuli contained main verbs (the retrieval site) that require animate subjects, like *complain*. They manipulated the features on a distractor noun phrase such that it either matched the subject (*the resident*) in animacy features (*the dangerous neighbor*) or did not match (*the dangerous warehouse*). The distractors matched targets on all other features required by the main verb (i.e., +NP, +SG). Their results suggested that the presence of an animate distractor NP resulted in longer reaction times at the main verb compared to when the distractor was inanimate.

They proposed that the mechanism operates as follows. At the main verb, the retrieval probe instantiates a search for an item stored in long-term memory bearing the features {+NP, +SG, +Anim}. Grammatically, this should result in retrieval of the ani-

Animacy	Sentence	**distractor** | *retrieval site*
+Anim	The worker was surprised that the resident who was living near **the dangerous neighbor** *was complaining* about the investigation.	
-Anim	The worker was surprised that the resident who was living near **the dangerous warehouse** *was complaining* about the investigation.	

Table 2.3: Critical comparison from Van Dyke and Lewis (2003). *Animacy* refers to the features of the distractor noun phrase.

mate head noun of the subject (*the resident*) because it is this item that controls number agreement on the verb. Because the retrieval mechanism is content-addressable, this search will activate the two NPs that match the probe in animacy features in the +Anim condition: the correct NP, *the resident*, and the structurally illicit but feature-matching NP, *the dangerous neighbor*. This process is represented schematically in (36a). The slowdown in reading times for the +Anim condition, then, is the result of competition between co-active NPs during retrieval. In the -Anim condition (36b), where the intervening NP *warehouse* is inanimate, RTs at the verbal region are relatively faster, because the distractor NP does not receive as much activation during the search process.

(36) Cue-based retrieval for conditions in Table 2.3

a. +Anim Condition

+Anim +Anim ?Anim

the resident **the neighbor** was *complaining*

b. -Anim Condition

+Anim ?Anim

the resident the warehouse was *complaining*

Furthermore, the retrieval mechanism is *direct access*, meaning that retrieval operations proceed in constant time. Support for this claim comes from studies inves-

Interveners	Sentence
1	It was **the scandal** that the celebrity relished __.
2	It was **the scandal** that the model believed the celebrity relished __.
3	It was **the scandal** that the model believed that the journalist reported that the celebrity relished __.

Table 2.4: Example item set from McElree et al. (2003), Experiment 1.

tigating the role of distance in establishing item-to-item dependencies. These studies find that increasing distance, whether linear or structural, between a retrieval-instantiating item and its dependent does not increase the amount of time it takes to complete a retrieval operation, but does affect overall accuracy. In other words, retrieving an item from memory does not require a serial search of each previously encountered item; this is a general property of memory that holds for both word lists and sentence contexts. For sentences, this has been determined using the Speed-Accuracy Tradeoff (SAT) response-signal procedure (McElree et al., 2003). In SAT studies, participants are trained to respond quickly and on cue, at varying time intervals, with an end-of-sentence acceptability judgement where the critical region (the position where a dependency must be resolved) is in sentence-final position. An example item set from McElree et al. (2003) is given in Table 2.4.

Based on SAT responses to these clefted object relative clauses, where each condition varies in the amount of intervening words and clause boundaries between the head of the relative clause and the gap site, McElree et al. (2003) determine that only the asymptotic accuracy of the characteristic SAT function, but not the slope or intercept, decreases as intervening material increases. McElree et al. take this to mean that accuracy is degraded due to partially feature-matching, interfering items in memory (*model*, *journalist*, and *celebrity* in the 3-Intervener condition, but retrieval speed (as determined by the slope and intercept of the SAT function) is not affected.

The claim that retrieval is direct access has consequences for the debate concerning the divide between long and short-term memory stores. Under McElree et al.'s model, the capacity of the focus of attention is assumed to be extremely limited, holding only one item (or data structure) at a time, and there is no differentiated short-term working memory (WM) store, where items can remain active in memory for a limited amount of time. Therefore, the focus of attention simply refers to activated long term memory, where items rise into and out of attention on a one-by-one basis. Computational models of memory, such as ACT-R (Anderson, 1993, 1996), instantiate these types of models without implementing a working memory buffer that can hold multiple items.

This departs from much of the research on linguistic segmentation discussed earlier in this section. Any account that makes a claim about the contents of the current segment being more active in memory, it must adopt a tripartite memory architecture that includes a short-term memory store. Under such accounts, not all items are directly accessible. The hypothesis advanced in §2.3.3 could adopt the assumption that there is no short-term memory store, because from a mechanistic standpoint, units in memory are assumed to be an epiphenomenon of features bound to items that encode properties of the context they occur in. On the other hand, I will refer to the process of reinstating previous contextual features at times. This could be spelled out by assuming that accessing a particular item makes it more likely that contextually nearby items will become active, or that the contents of a segment can be reactivated as a whole. Throughout the remainder of the dissertation, I remain neutral about whether the structure of memory is bi- or tripartite, as either structure could be potentially be made consistent with the account I propose here.

2.3.2 The Temporal Context Model

In order to account for the idea that item membership in a particular linguistic segment is a property that can be encoded incrementally along with other item-level features, the hypothesis advanced here adopts the Temporal Context Model (TCM) of memory (Howard & Kahana, 2002). This section first introduces the basic assumptions of the TCM and reviews evidence for a notion of temporal context from studies on list memory. It then turns to investigations of how TCM relates to sentence memory.

2.3.2.1 Temporal Context in List Memory

The TCM predicates on the idea that the formation of item-level memories does not occur in isolation, but is rather situated in a particular spatiotemporal context, such that items are associated with a representation of the context they were encoded in. The fact that properties of the external context affect memory has been long acknowledged by memory researchers. For example, in word list recall studies, holding external environment constant during study and test phases yields a benefit to memory, whereas testing in a different environment than the study location reduces recall accuracy (see Godden & Baddeley, 1975, for example).

In addition to properties of the physical environment, investigations of temporal context suggests that *internal context*, or a representation of the pattern of activity in the cognitive system, is also associated with item encodings. The TCM presents a mathematical framework for capturing this fact, representing the context as a slowly-evolving vector that is both associated with item encodings, as well as changed by them (Howard & Kahana, 2002). In this way, each item is marked with a featural representation analogous to an address or timestamp, which blends information about the cognitive system's internal state with information about pre-existing contextual associations for each of the studied items. Thus, changes in the contextual representation are also driven by the properties of recently encountered items.

The equation for contextual update is given in (37).

(37) $c_{\mathrm{i}} = \rho_{\mathrm{i}} c_{\mathrm{i\text{-}1}} + \beta c^{IN}$ (Howard & Kahana, 2002)

Assuming that the context vector is incrementally updated at each timestep, the context at the current timestep, i, is derived from the contextual state immediately preceding the presentation of i, $c_{\text{i-1}}$, and the vector c^{IN}, which contains temporal and preexisting elements of the new context. β is a parameter that determines the strength of new contextual information, and ρ_{i} downweights $c_{\text{i-1}}$ such that the overall level of contextual activation remains constant (see Howard and Kahana (2002) and Polyn, Norman, and Kahana (2009) for details).

The model proposes bidirectional associations between items and their encoding context, and between contexts and the items encoded in them. Therefore, the retrieval of an item partially reinstates the context that item was encoded in, and retrieval of a contextual state reinstates items encoded at that "timestamp" or items encoded nearby. This is due to the evolution of the context vector, which serves to relate the organization of items in memory to the passage of time: two items encoded in (perceived) close temporal or positional proximity will be more contextually similar to one another than the same two items would be if they were not temporally or spatially adjacent. In other words, items encoded in perceived temporal proximity are associated with similar contextual states. Polyn and Cutler (2017) analogize the contextual retrieval cue to a spotlight: once an item is retrieved, its retrieved context is akin to a spotlight centered on that item's contextual state, whereby neighboring items at the time of encoding will also be partially illuminated.

Early work on the role of temporal context dealt with *recency* and *temporal contiguity* effects in memory. The recency effect refers to the well-established recall benefit for the most recent list item. The TCM accounts for this effect because the contextual state at the beginning of the test phase of an experiment will be most similar to the contextual state of the most recently studied item. The contiguity effect refers to the fact that recall of a particular item is very likely to evoke recall of items that occurred in close temporal proximity during encoding. Studies measuring conditional response probabilities show that during free recall from unstructured word lists, recall of a word w_{i} is most likely to result in recall of words $w_{\text{i+/-1}}$ (see Healey, Long, and Kahana, 2019, for a recent review). Contiguity effects also arise in recall latencies. Latency between recalls of proximal list items is significantly shorter than more distant ones (Kahana, Howard, & Polyn, 2008). Greater ease and likelihood of recall from contiguous positions follows directly from the spotlight analogy above.

Additionally, these studies report a robust forward-asymmetry: that recall of word $w_{\text{i+1}}$ is much more likely than $w_{\text{i-1}}$. The TCM accounts for the forward asymmetry through of the integration of pre-existing contextual associations with the timestamp of a given item. If an item brings along some set of contextual associations during encoding, which are used to update the current state of the context at t_{i}, then the item encoded at $t_{\text{i+1}}$ will be contextually more similar to item i than the item encoded at $t_{\text{i-1}}$, which is not associated with item i's pre-existing context (Sederberg, Howard, & Kahana, 2008). A notable property of both recency and contiguity is that they are time-scale invariant. That is, the contiguity effect is robust to the amount of absolute inter-item delay between list items, as well as the retention interval between the last list item studied and the test phase. This suggests that it is the relative interval between items that determines their contextual likeness.

In this way, TCM offers a natural explanation for the formation of groups: an item following a relatively longer delay than the rate at which a sequence of previous items was presented is likely to be accompanied by greater contextual distinctiveness. The temporal structure of items is thus an important organizing property of memory. Correspondingly, contiguity effects obtain across lists as well: when participants study multiple lists, if they recall an item not contained within the same list, they are more likely to transition to a temporally adjacent list (Howard, Youker, & Venkatadass, 2008;

Unsworth, 2008). This across-list contiguity effect also shows the forward asymmetry. This suggests that participants retain and utilize information about larger group membership in memory. Furthermore, this suggests that temporal context is equipped to deal with hierarchical representations, i.e. those that capture inter-group relationships, as chunking or hierarchical associative models are (see Kahana et al. (2008) and Farrell (2012) for further discussion).

Under the TCM, group membership is accounted for as follows. TCM supposes that the representation of temporal context may serve to partition experiences into distinct episodes[10]. Polyn and Cutler (2017) suggest that a significant "disruption" (construed broadly) can lead to a sudden shift in the contextual representation (see the following section for a related proposal by Wagers, 2008). They relate this to studies that have induced disruptions via shifts in task, category, or spatial location, and using words in narratives that indicate the passage of time (see references in Polyn & Cutler, 2017). Similarly, Polyn et al. (2009) suggest the following:

> "...the principle of clustering by isolation states that a sudden shift in context (caused by a disruptive cognitive event) can isolate a set of items from the items studied prior to the disruptive event. This causes the isolated items to cluster together in the recall sequence (relative to a condition without such a disruptive event)."

The sensitivity of contextual representations to group membership has various consequences. Importantly, participants are able to access and manipulate contexts independently from one another. In experiments where participants are presented with multiple lists separated by a delay, they are able to effectively use list membership as a cue to restrict memory search processes. For example, participants can accurately judge (i) the modality a particular item was presented in and recall other items presented in the same modality, (ii) the frequency of items across different list

[10]This is reminiscent of what Zwaan and Radvansky (1998) suggest for temporal and spatial discontinuities under Situation Model Theory.

contexts, (iii) the specific position in which a particular item appeared, and (iv) which items appeared in specific lists (Anderson & Bower, 1974; Hintzman & Block, 1971, 1973; Hintzman, Block, & Summers, 1973; Light, Stansbury, Rubin, & Linde, 1973). These judgments are performed with a high degree of accuracy even without specific instruction to attend to contextual cues during study.

Before concluding this section, we note an important difference between adopting a notion of temporal context and adopting a notion of *chunks* in memory. Chunking assumes that smaller units of information are re-coded into higher-order representations such that more content can occupy the short-term store. Modern models of linguistic memory, like cue-based retrieval, often rely on the premise that item-level features are accessible at a distance. This necessitates that low-level features are retained, even after the formation of higher-order hierarchical structures (see C. Andrews, 2021, for a detailed discussion of this point). Because TCM in principle allows multiple, consecutive chunks to be encoded as contextually similar, there is a possible role for chunking under the current framework, although many accounts of TCM discuss chunking models as if they are mutually exclusive (Kahana et al., 2008). In the current sentential implementation, I assume that a single encoding context may contain multiple linguistic “chunks” (i.e., units of structure). In other words, the proposal that certain linguistic boundaries can induce a shift in context is not meant to replace the need for hierarchical representations in language. Because the TCM allows for a notion of linguistic segmentation without re-coding, I do not explicitly discuss the role of chunking in the remainder of the dissertation.

2.3.2.2 Temporal Context in Sentence Memory

The integration of TCM into sentence processing was first proposed by Wagers (2008) to account for the fact that interference effects in language are sometimes sensitive to clause membership. The proposal was that clause membership information is tracked along with item-level properties of constituents as a context encoding. Wagers' proposal for marking clause boundaries mirrors the suggestion about encoding distinct episodes discussed in the previous section. Specifically, when the parser crosses a clausal boundary, the new contextual state in (37) is argued to be shifted by a randomly generated c^{IN}, in order to contextually distinguish incoming input from preceding content.

While Wagers (2008) considers the utility of such shifts at clausal boundaries or syntactic domains, we suggest here that the advantage of an elaborate contextual representation is that it is ripe for tracking multiple levels of linguistic structure, of potentially varying sizes. This is desirable given Tanenhaus and Carroll's (1975) suggestion that there is no fixed boundary type sufficient to enforce segmentation, and given the varying evidence for segmentation effects across different levels of linguistic representation reviewed in §2.1. Crucially, we will assume here that the same context shifting process can occur at prosodic boundaries. The question of what types of boundaries provide a sufficient cue to trigger a context shift is an important one, but one that the current dissertation will not be able to definitively address, because the constructions utilized here align with syntactic, prosodic, and discourse boundaries. Nevertheless, the results reported here ultimately show that prosodic boundaries play a special role.

In joint work with Stephanie Rich on the role of temporal context in sentence memory, we uncover two findings of note. First, contextual cues can be used to reactivate the contents of a particular clause. Second, prosodic and clausal boundaries serve to reduce encoding interference among similar items. I discuss each of these findings in turn.

In a sentence recall study, we presented participants with multi-clausal lists each containing four conjuncts, as in (38). Participants read these sentences using chunk-

Condition	Question
NoQ	–
Clause2-Q	Who did the miner love?
Clause3-Q	Who did the enemy love?

Table 2.5: Example item set from Rich (2024), Experiment 4.

by-chunk cumulative self-paced reading. The design manipulated the presence of a sentence-final comprehension question, as in Table 2.5, which participants provided free responses to. Questions served as a cue to reactivate either the second or third clause. In these conditions, participants were always prompted with an object *wh*-question; that is, they were explicitly prompted with the subject of the target clause and were asked to provide the object. A third baseline condition was not followed by a question. Participants were then asked to recall the sentence they read on the next screen.

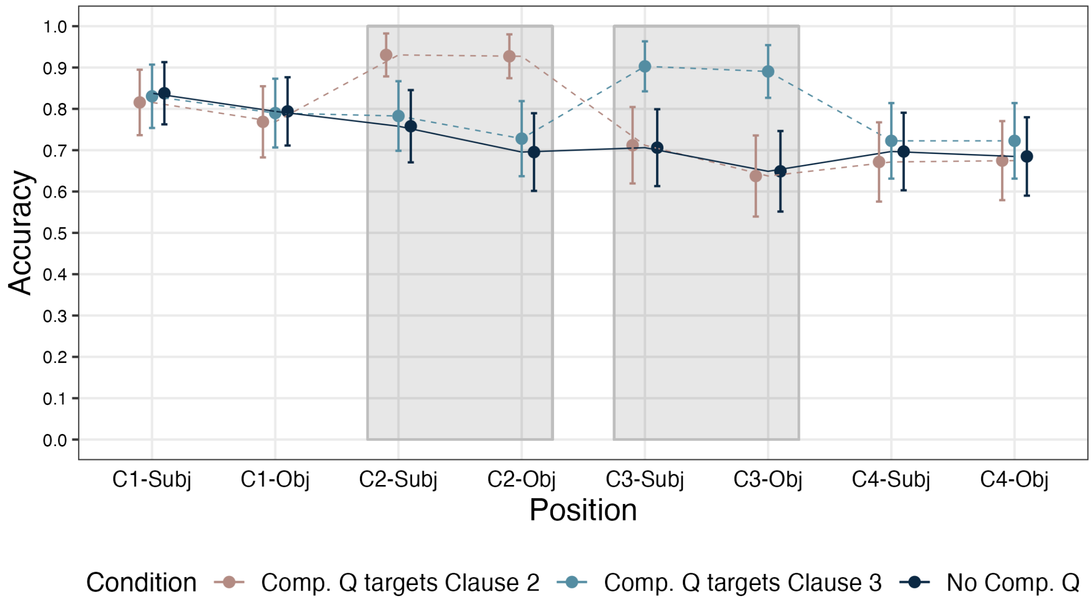
(38) Exposure sentence from Rich (2024)
The guest loved the voter, the miner loved the guide, the enemy loved the groom, and the boxer loved the artist.

We hypothesized that if the effect of contextual reactivation is strictly clause-bounded, we should observe a recall benefit only for the targeted clause. If instead a clause-level contiguity effect obtained, we expected a benefit for the subsequent clause when Clause-2 was targeted and a smaller benefit for the previous clause when Clause-3 was targeted, per the forward asymmetry effect observed for contiguous items and chunks. Alternatively, if linguistic boundaries did not serve to shift the contextual representation, we expected to find a temporal contiguity effect such that the position immediately following the target clause alone (either the subject of clause 3 or the subject of clause 4) should show a recall benefit following successful recall of the cued object.

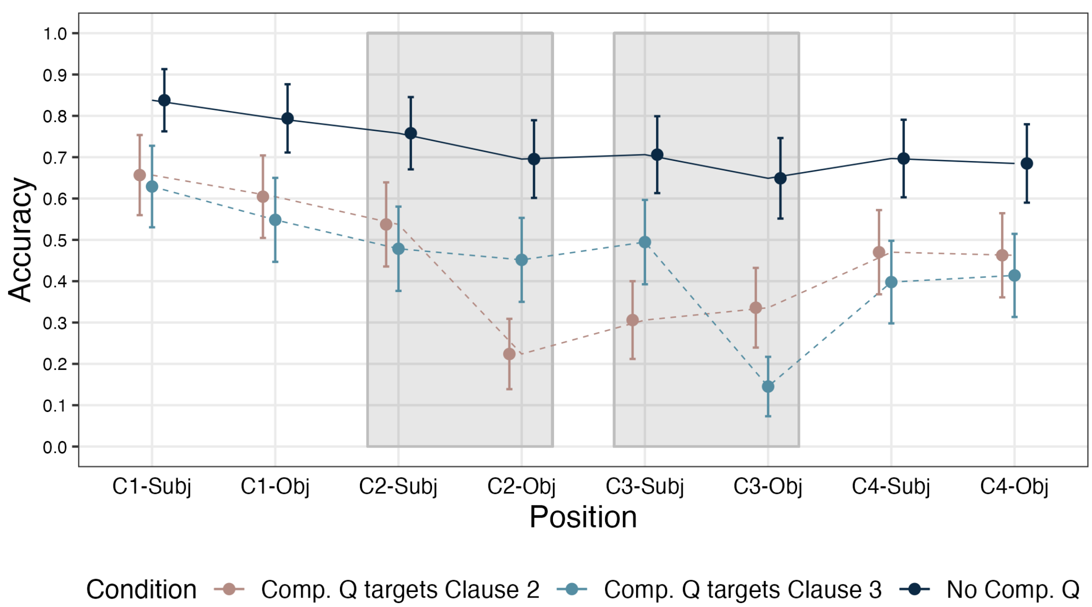
We found that for trials where participants correctly responded to the comprehension question, only the targeted clause showed a recall benefit (Figure 2.1a). For trials where participants incorrectly responded, recall accuracy for the targeted clause showed a significant decrement, and in addition, recall of the following item (the subject of the subsequent chunk, labeled C3-Subj) was also negatively impacted (Figure 2.1b). Recall that successful recall of an item should lead to a high probability of successful recall for the subsequent item as well. Therefore, unsuccessful recall of an item in list position i provides a degraded cue to position $i + 1$. Similarly, the pattern of results on incorrect trials suggested that failure to reactivate the cued object led to disruption in participants' contextual representation, leading to degraded accuracy for the subsequent position as well. This decrement for the contiguous subject position surfaced only for the Clause2-Q condition, not the Clause3-Q condition, where the subsequent subject position belong to the final clause. This suggested a recency benefit for the final clause as a whole, rather than only the final element (the object of clause 4).

On the whole, the results suggest that clause boundaries serve to partition linguistic segments in such a way that they are able to be reinstated to the exclusion of other sentence content later on. This conclusion is consistent with work by Jarvella (1973), which showed that in multi-clausal discourses, memory for co-referential clauses (where the second clause contained an element co-referential with an earlier one) shows higher recall and recognition accuracy than non-co-referential clauses. Together, these results provide support for the idea that each linguistic segment is associated with a distinct contextual representation in memory, and that these segments may be reactivated for later processing purposes.

In another recognition memory study reported by Rich (2024), we aimed to determine whether prosodic and clausal boundaries can serve to reduce potential encoding



(a) Recall accuracy by position for correct question trials.



(b) Recall accuracy by position for incorrect question trials.

Figure 2.1: Recall accuracy by position from Rich (2024), Experiment 4.

	MATCH	MISMATCH
WITHIN	Before the butcher and the landlord could, **the scientist** and **the researcher** solved the problem.	Before the butcher and the landlord could, **the researcher** and **the scientist** solved the problem.
ACROSS	Before the butcher and **the scientist** could, **the researcher** and the landlord solved the problem.	Before the butcher and **the researcher** could, **the scientist** and the landlord solved the problem.

Table 2.6: Example item set from Rich (2024), Experiment 5. Semantically similar NPs are bolded.

interference between conceptually similar items (those with highly related `word2vec` (Mikolov, Chen, Corrado, & Dean, 2013) scores). In particular, we tested temporal order memory of similar items when they either occupied the same clause (as conjoined nouns in subject position) or spanned a clause boundary. Our design crossed Match (MATCH, MISMATCH), whether the temporal order of two nouns was identical or reversed, with Position (WITHIN, ACROSS), as in Table 2.6.

We predicted that if clausal/prosodic boundaries reduce potential encoding interference among similar items by shifting the temporal context representation at linguistic boundaries, the discriminability of temporal order mismatches should be greater in the ACROSS condition compared to the WITHIN condition. Our results supported this conclusion (but see Rich, 2024, for alternative explanations).

Some previous list memory work suggests that across group temporal order memory is degraded compared to within group ordering (Pu, Kong, Ranganath, & Melloni, 2022a). However, other studies suggest that this across-group penalty is reversed when similar items span group boundaries, perhaps because they serve to provide a link between groups (Gurguryan, Dutemple, & Sheldon, 2021). Our results are consistent with the latter view, because our items included conceptually similar nouns across clause boundaries. It remains to be seen whether the linguistic relationship between subordinate and main clauses, like the structures we used in our experiment, are sufficient to lead to the same boost for temporal order memory across boundaries, or whether the results were influenced by our similarity manipulation in particular. The TCM should predict that linguistic boundaries reduce encoding interference more generally, not just for similar items or temporal order memory.

These studies have established some preliminary evidence for the role of temporal context in offline sentence memory. The Context-Sensitive Encoding Hypothesis advanced in the current dissertation will attempt to extend this work by importing some basic principles of TCM into the domain of real-time sentence processing, following Wagers (2008).

2.3.3 The Context-Sensitive Encoding Hypothesis

I propose the Context-Sensitive Encoding (CSE) Hypothesis in (39), which advances the idea that linguistic segmentation proceeds via a temporal context mechanism that marks significant boundaries during encoding. The definition of CSE in (39) proposes that unambiguous cues to sentence-medial prosodic boundaries[11] (like commas or other punctuation in the orthography) can provide a particularly salient and immediate signal to shift the contextual representation associated with incoming input. CSE assumes that context shifts should occur probabilistically in response to both implicit and overt prosodic boundaries. Once the boundaries of a medial segment have been marked, CSE proposes that the parser may treat contextually similar content bounded by context shifts as a distinct *encoding context* in memory. Mechanistically, the TCM assumes that the formation of these groupings is epiphenomenal; therefore, I adopt this terminology for the sake of simplicity. However, this process does not preclude that the resulting linguistic segments are independently represented as units in mem-

[11]This is not to say that other types of linguistic boundaries could not influence context shifting. I begin by considering the role of prosodic boundaries because they evidence clear segmentation effects in memory (Jarvella, 1979; Simpson, 2016), and because implicit prosodic boundaries show promise in explaining appositive bypassing and discounting effects. See §3.

ory within a hierarchical structure. In fact, one way to conceive of context encodings for sentences is that they have a dual role: (i) to mark the position of boundaries online and (ii) to bind items to their positions in a syntactic representation, such that information about nearby syntactic relationships and larger syntactic domains is also encoded. Such a view could capture how a representation of the global linguistic structure is built online. Crucially, the experiments reported here will show that different linguistic processes may make use of such contextual information in different ways.

(39) **Context-Sensitive Encoding** (CSE) **Hypothesis**: Prosodic boundaries shift the *temporal context vector* during encoding, such that prosodic phrases partition syntactic content in memory into distinct *encoding contexts* (EC), and the content within an EC shares similar contextual features. These ECs are built up incrementally and may be reactivated on a by-context basis at a later stage, using positional cues. Thus, segmented sentences incur less processing load than non-segmented ones, due to decreased contextual interference during encoding and retrieval.

Given a sentence containing multiple encoding contexts, CSE predicts two consequences. The first is that a sentence with greater prosodic segmentation should also benefit from finer structuring in memory relative to an otherwise identical sentence with fewer sentence-internal boundaries. This follows from the finding that *retrieval set size* influences interference effects in experiments that vary the length of different list contexts. For example, Oberauer (2002) presented two lists of digits during a study phase, then presented participants with a visual cue marking an active list, which was to be immediately used for further processing. Participants were instructed to apply an arithmetic operation to each item in the active list. In the meantime, participants were also asked to remember the passive list. After completing the arithmetic task, participants recalled both the active and passive lists. The results showed that arithmetic latencies were modulated only by the set size of the active list, not the passive one. Oberauer's interpretation of the results argued for a tripartite memory structure, where both the active and passive lists can be held in separate short-term memory stores, but only the active list remains in a capacity-limited store that affects working memory load.

An alternative interpretation of these results under TCM is that the active and passive lists were encoded with distinct contextual features. That is, the boundary between lists triggered a context shift, and the visual prompt to access the active list acted as a contextual retrieval cue. Therefore, when the active context was accessed, contextual interference based on the size of that context alone arose during the arithmetic task. Afterwards, a contextual retrieval cue to the passive list could still be used to recall the contents of that list without it interfering with the active task.

At the sentence level, CSE assumes that as more lexical and/or syntactic content is added to a particular context during encoding, the contextual interference engendered by the contents of that context increases. Thus, a more segmented sentence should result in less contextual interference during encoding. This is exemplified via the contrast between (40) and (41), where the contents of each hypothesized encoding context is boxed in a different color (distinctions between encoding contexts will be represented as such throughout the rest of the dissertation).

(40) Sentence-medial context shift

Since April bought apples from the market, Ethel baked a homemade pie.

(41) No sentence-medial context shift

Ethel baked a homemade pie since April bought apples from the market.

Similarly during recall or recognition, accessing a portion of a sentence segmented into distinct contexts should be less costly than recall or recognition of a non-segmented sentence. This is because in the segmented case, contextual retrieval cues to each segment can be utilized in a more targeted manner, thus improving the precision of retrieval and reducing associated load due to contextual interference. CSE therefore predicts that contextual interference should be reduced for (40) compared to (41).

A potential consequence of CSE during online processing is that the effect of contextual interference may arise following item-level retrieval processes as well. This motivates the proposal of the Context-Sensitive Retrieval (CSR) Hypothesis in (42).

(42) **Context-Sensitive Retrieval** (CSR): the cue-based retrieval mechanism can use the context vector as a cue, because retrieval of an item reinstates that item's encoding context. Therefore, retrieval of an item from a larger encoding context (i.e., one that contains more lexical/syntactic content) incurs greater contextual interference than that of a smaller encoding context.

CSR predicts that the effect of contextual interference should appear immediately following the retrieval of a target item. Specifically, it suggests that the resolution of an item-to-item dependency, like anaphora, should induce greater interference due to contextual similarity in a non-segmented sentence like (44) compared to a segmented one like (43). In (44), this should result in longer reaction or reading times on the critical words *two of them*, which cues the dependency. This prediction follows from the retrieval set size effect as well: if the content preceding time ① is encoded with similar contextual features, and only contextually similar content has the potential to receive some amount of activation following retrieval at time ②, then the second encoding context in (43) should not contribute further interference.

One challenge with searching for effects of CSE using structures like in (40)-(41), and for CSR using (43)-(44), is that these pairs differ in terms of word and clause order. In addition, (40) and (43) involve a syntactic expectation for a main clause during the encoding of the initial context, whereas (41), for example, does not lead to an expectation for a subordinate clause. Without holding these differences constant, it's difficult to probe the memory processes at work.

(43) 1 2

Since April bought **four apples**, Ethel used **two of them**.

1 : *apples,* → context shift

2 : *two of them* → retrieve *apples*, reinstate its encoding context

(44) 1

Ethel used **two apples** since April bought **four of them**.

1 : *four of them* → retrieve *apples*, reinstate its encoding context

The remainder of this dissertation probes CSE/CSR using the structures in Table 2.7. The contrast between appositive and restrictive relative clauses, which is the focus of Chapter 3, offers a useful starting point. Although these structures differ along other important linguistic dimensions, discussed at length in §3, this contrast holds word and clause order constant, as well as the number of clauses contained within each sentence. Chapters 4 and 5 focus on the contrast between prosodically segmented vs. prosodically integrated coordination, which offers another set of minimally different pairs.

CSE lends itself to an additional prediction in the case of sentence-medial appositives, where main clause material is separated: that a previous context can be reinstated in order to add content to it, using a positional retrieval cue to the relevant context. This assumption follows from work on list memory by Hupbach, Gomez, Hardt, and Nadel (2007) and Sederberg et al. (2011), which suggests that when provided with a cue to a previously studied list, participants can reinstate that list's encoding context for the purpose of updating it with additional items. Evidence for this process comes from Hupbach et al.'s (2007) list learning paradigm, where participants stud-

Chapter	Segmented Structures	Non-segmented Structures
Chapter 3 (Exp. 1-3)	**Appositive Relative Clause** The cat, who loves cheese, was taking a nap.	**Restrictive Relative Clause** The cat that loves cheese was taking a nap.
Chapter 4 (Exp. 4)	**Isolated Coordination** The cat ate not only the cheese, but also the eggs.	**Integrated Coordination** The cat ate the cheese and also the eggs.
Chapters 4-5 (Exp. 5-8)	**Isolated Coordination** The cat, as well as the dog, were taking a nap.	**Integrated Coordination** The cat {and, as well as} the dog were taking a nap.

Table 2.7: Example structures probing CSE/CSR in the remainder of the dissertation.

ied items from two lists, A and B, with a 24 hour interval in between. One group of participants was provided with a reminder of learning list A prior to studying the list B items. These participants were guided by the same experimenter to study list B in the same room as list A. In the control condition, participants were not provided with a reminder and learned list B in a different room with a different experimenter. The results evidenced an asymmetric intrusion effect in the reminder condition, where when participants were asked to recall list A at a delay, a significant number of list B items intruded, but when asked to recall list B, list A items did not intrude. Hupbach et al. (2007) suggested that the reminder of list A rendered memory for this list more malleable, such that list B items were used to update list A rather than being encoded as entirely contextually distinct.

Sederberg et al. (2011) contextualize these results within TCM by assuming that the list A items are linked to a list A context, and list B items are linked to a list B context in the no-reminder condition. Given a reminder of list A, this reinstates the list A context and causes list B items to be linked to the list A context and to also be uniquely associated with the list B context. Thus, the asymmetric intrusion effect arises from the fact that the list A context is linked to items from lists A and B, whereas the list B context is linked to list B items alone. Sederberg et al. report the results of TCM simulations that support this as a viable mechanistic explanation. I term the sentence-level analogue of this process Reinstantiation (45).

(45) **Reinstantiation** (to be revised): The contents of an earlier encoding context can be reinstated using positional cues in order to add additional material to a previously incomplete linguistic segment (c.f. Sederberg et al., 2011).

In principle, CSE predicts that reinstatement of a previous contextual state should be possible during the processing of a sentence, not only during recognition or recall tasks. Recall Grosz and Sidner's (1986) suggestion that in sentences containing interruptions such as (46), the first and third segments form a single discourse unit (D1: *John left the groceries and I put them away*). Under a CSE-based model where Reinstantiation is possible, this is schematically represented in (46b). As discussed in §2.1.3.2, one issue with positing such an online process for Grosz and Sidner's example is determining what the relevant cue to reinstate the initial context at the time of encoding would be. In theory, the contextual representation could just as easily look like (46a) following initial encoding, which reflects a pure effect of boundaries under CSE where the first and third segments are contextually distinct. Perhaps the contextual state associated with the final segment can then be updated to arrive at (46b) in the sentence-final representation, once the discourse relationships between segments have been established.

(46) a. CSE

$\boxed{\text{John left the groceries}}$ – $\boxed{\text{stop that you kids}}$ – $\boxed{\text{and I put them away.}}$

b. CSE + Reinstantiation

$\boxed{\text{John left the groceries}}$ – $\boxed{\text{stop that you kids}}$ – $\boxed{\text{and I put them away.}}$

During the processing of appositive relative clauses, however, where part of the main clause subject is prosodically set off, an expectation for Reinstantiation may be determined in advance of encoding the final segment. If sentence-medial punctuation (commas, parentheses, or em dashes) act as cues to trigger context shifts and a top-down syntactic expectation to continue a previously incomplete context (i.e., the knowledge that the subject must form a dependency with an upcoming verb) leads to reactivation of that context's most recent contextual state, then the contextual state associated with the first segment should be relatively more available in advance of encountering content in the third segment. This contrasts with what a pure effect of prosodic boundaries during retrieval might predict, as per CSR, for example. The incremental timecourse of Reinstantiation for ARC structures is provided in (47).

(47) Timecourse of CSE + Reinstantiation

1 2 3

The cat, who loved cheese, **was** taking a nap.

1 : *cat,* → context shift

2 : *cheese,* → context shift, reinstate initial context

3 : *was* → retrieve subject (*the cat*), with initial context more active

To preview the conclusions from later chapters, the work reported here provides convincing evidence for CSE and Reinstantiation processes, but suggests that the effect of CSR is only observable under particular syntactic conditions. Specifically, evidence for these processes emerges given the presence of a prosodically isolated linguistic segment bounded on both sides by hypothesized context shifts, along with a top-down expectation for a subject-verb dependency. Therefore, the dissertation advances an account of the online processing of ARCs that resorts not to their discourse status, but rather their prosodically-driven segmentation in memory.

Finally, I return to a discussion of Visibility. Recall the conclusion of the recency-based view from §2.2.2: that attachment sites in the current prosodic phrase are most visible. This particular conception of Visibility has a clear explanation under the TCM. Recall that the TCM proposes that list recall studies often result in recall of the most recently studied item first because the contextual state during the test phase most closely matches the contextual state of the last item. Similarly, under the assumption that prosodic boundaries delineate encoding contexts, the current or most recent prosodic phrase should most closely match the contextual state of the incoming, to-be-attached phrase. Therefore, all else equal, the parser should prefer a contextually-like attachment site. In the event that a previous contextual state is accessed (e.g., because there is no attachment site available within the current encoding context), there may be some cost associated with accessing its contents[12]. An updated definition of Visibility that relies on a temporal context mechanism is provided in (48).

(48) **CSE-Visibility**: The contents of the current encoding context (i.e., the current prosodic phrase) are most accessible, because they most closely match the current contextual state. Retrieving the contents of an earlier encoding context requires reinstatement of an earlier contextual state.

I return to a discussion of this mechanism in Chapter 4 (§4.1.1).

[12]This aligns with work on list memory that observes switch costs for loading the contents of a previous list into working memory (see Oberauer, 2005, 2009, for example). I thank Brian Dillon for drawing my attention to this connection.

Chapter 3

Appositives in Memory

Chapter 3 investigates the status of appositive relative clauses (ARCs) in memory, like in (1a).

(1) a. The sailors, who race Santa Cruz 27s[13] on the bay, are very experienced.

b. The sailors that race Santa Cruz 27s on the bay are very experienced.

ARCs offer an especially useful starting point for the current investigation. Sentences containing ARCs are comprised of discrete prosodic and pragmatic sub-sentential units, and it has been well-established across a series of recent experimental studies that appositive content contributes less to later processes than integrated structures like the restrictive relative clause (RRC) in (1b), an effect which Duff et al. (2023) term *discounting*. One explanation that has been put forward to explain this observation is that they hold a degraded status in memory relative to other sentence content (Dillon et al., 2017); however, little empirical evidence directly supports this explanation. Moreover, appositives are complex constructions with a constellation of interesting linguistic properties; they are argued to be independent from the main clause (their hosts) at syntactic, prosodic, semantic, and discourse levels of representation. Despite

[13] A type of (fast and fun) sailboat designed in Santa Cruz in the 1970s.

agreement that they must be special in some sense, there is little consensus in the theoretical literature with respect to the appropriate level of representation at which to situate their independence. Similarly, in the psycholinguistic literature on appositives, the source of discounting is actively debated. Previous investigations have considered the potential roles of syntactic, prosodic, and discourse independence of various sorts, but ultimately reach different conclusions.

This chapter discusses a suite of linguistic properties associated with appositives before turning to their unique processing profile. It puts forth a proposal based on the Context-Sensitive Encoding and Reinstantiation Hypotheses, which accounts for the tendency to bypass ARC content during sentence comprehension. In addition, it explores whether retrieval operations underlying item-to-item dependencies can benefit from the segmentation of linguistic units during on-line processing, per Context-Sensitive Retrieval (CSR). While appositives may indeed be by-passed once their processing is complete, Experiments 2-3 show that their contents remain equally accessible to the cue-based retrieval mechanism. This is not in line with CSR, but it also refutes theories that rely on the subordinate status of ARC content. On this basis, the current chapter proposes that the prosodic boundaries demarcating appositive units drive contextual distinctiveness between the contents of ARCs and their hosts during encoding, and that the resulting contextual partitions along with the top-down expectation for a main verb allow for the reactivation of previous incomplete contexts during incremental comprehension. The chapter ultimately argues that their processing behavior can be derived without reliance on their subordinate discourse status, and motivates why such an approach is desirable. At a higher level, it aims to shed light on how units of linguistic structure map onto units in memory, and how different levels of linguistic representation interact with one another in memory.

3.1 Properties of Appositives

3.1.1 Linguistic Properties

Appositive relative clauses (1a) are a type of parenthetical structure, which are prosodically isolated from their containing sentences and are argued to contribute *not-at-issue* content (Potts, 2005). In contrast, restrictive relative clauses (1b) contribute *at-issue* entailments to the sentences that contain them. Although appositives introduce new, entailed content (Schlenker, 2010; Syrett & Koev, 2015), they are thought to contribute secondary information, which need not be directly relevant to the main point of the sentence as a whole. Thus, Chierchia and McConnell-Ginet (1990) suggest that sentences like in (1a) contain a main assertion which they term the foreground (*the sailors are very experienced*) and a secondary point termed the background (*the sailors race Santa Cruz 27s*).

3.1.1.1 Semantic & Pragmatic Properties

Potts (2005) posits that the not-at-issue status of appositives results from their semantic independence; formally, they are argued to compose independently from host clause material in the semantic representation. Evidence for semantic independence of appositives comes from the fact that they are inaccessible for binding from the host clause, and that they can project past negation, questions, and propositional attitude verbs, like other types of not-at-issue meaning (e.g., presuppositions). The question in (2), for example, does not ask whether the appositive content is true, but rather asserts the appositive content independent of the main interrogative content. In (3a), it's clear that the content of the appositive is speaker-oriented, and not necessarily a belief that Libby holds; in (3b) where the same content is conjoined and at-issue, the resulting sentence is infelicitous, suggesting that the ARC content in (3a) is not interpreted within the scope of the verb *believed*. Additionally, negative polarity items (like *any*) within an ARC are not interpreted in the scope of main clause negation, as evidenced by the contrast between (4a) and (4b) (Burton-Roberts, 1999).

(2) Is the Santa Cruz 27, which was designed by Bill Lee, fun to race?

(3) a. Libby believed that the students, who had just begun learning to sail, were very experienced.

b. #Libby believed that the students had just begun learning to sail and were very experienced.

(4) a. None of the authors who had any imagination reviewed the book positively.

b. *None of the authors, who had any imagination, reviewed the book positively.

Further evidence for semantic independence comes from the observation that ARCs do not restrict reference, whereas RRCs do. (1b), for example, picks out a particular subgroup of sailors (those that race Santa Cruz 27s). (1a), on the other hand, does not pick out a subgroup of sailors relative to some possible larger set, but rather comments on a property of the entire set.

Despite this, it is not the case that appositives must always be completely independent from their hosts. For example, AnderBois, Brasoveanu, and Henderson (2010) observe that presuppositions, ellipsis, and pronominal anaphora are permitted to "cross" appositive boundaries, targeting ARC-internal content. In (5), the antecedent of the main clause ellipsis site is contained within an ARC. (6) exemplifies the ability to establish reference between a pronoun and its antecedent in both the Main → ARC direction (*John-his car*) as well as the ARC → Main direction (*a woman-her*). Thus, it is possible to resolve syntactic and semantic dependencies across appositive boundaries. This does not, however, preclude the fact that there may be processing differences between dependency resolution in each of these cases; this is further discussed in Experiments 2-3 of this chapter.

(5) Melinda, who won three games of tennis, lost because Betty won six <~~games of tennis~~>.

(6) John$_i$, who nearly killed a woman$_j$ with his car$_i$, visited her$_j$ in the hospital.

Because of this, others have situated their independence not in terms of their semantic composition, but instead, their pragmatic independence. As alluded to above, for instance, the appositive in (2) contributes an independent speech act, because its declarative contribution is distinct from the interrogative force of the host content. This is especially apparent in a situation where the speaker is questioning one interlocutor about the main clause content and addressing the appositive content to a different interlocutor, like in (7).

(7) Context: *A has never sailed a Santa Cruz 27 before, but knows what kind of sailboat it is. A knows that B often races Santa Cruz 27s.*
A to B: Is the Santa Cruz 27,
A to C: which is an ultra-light displacement sailboat (by the way),
A to B: fun to race?

In fact, given the context in example (7), it would be odd for A to direct the appositive content at B knowing that B already has knowledge of what kind of boat the SC27 is. The observation that it is infelicitous for appositive content to state given or shared knowledge has been termed the *anti-backgrounding requirement* by Potts (2005) (but see Schlenker (2010) and Syrett and Koev (2015) for a discussion of exceptions to this rule).

Evidence for their independent discourse status is also supported by the fact that it is difficult to directly reject appositives, unlike RRCs. In (8a), a rejection like *No,* *he's not* is much more likely to target main clause content. ARCs may, however, be indirectly rejected via an exclamative like *Hey, wait a minute!*, as in (8b).

(8) A: Noodle, who's a cat, is a real menace.

a. B: No, he's not {a menace, #a cat}.

b. B: Hey, wait a minute! {He's not a menace!, He's a dog!}

I note, however, that the infelicity of directly rejecting the ARC in (8a) may simply be due to the fact that the appositive there is generally difficult to disagree with and unlikely to be in need of correction. In (9), where the ARC contributes a speaker-oriented attitude (and the ellipsis site trigger matches the tense of the ARC but not the main clause), it seems easier to directly reject this content. Similarly in (10), where the content of the appositive is more easily confusable (i.e., two similar sailboats), it seems more plausible that speaker B could offer a correction targeting the ARC.

(9) A: Noodle, who's a real menace, was chewing on the plants.

B: No, he's not! (He's actually really sweet).

(10) A: Libby's boat, which is a Santa Cruz 27, is fun to race.

B: Actually, no it's not. (It's a Moore 24[14].)

Moreover, the tendency for ellipses like in (9)-(10) to prefer to target the main clause may be a consequence of recency of the main clause VP, rather than the discourse status of the ARC. This explanation evokes general processing constraints, rather than a linguistic property of ARCs. In accord with this, Syrett and Koev (2015) present experimental evidence exhibiting that comprehenders are much more likely to target a sentence-final ARC (11a) via a direct rejection than a sentence-medial one (11b). They, however, attribute this to the fact that sentence-final ARCs are more easily able to achieve at-issue status. This aligns with Jasinskaja's (2016) proposal that

[14] Another 1970s-era Santa Cruz sailboat.

the at-issue status of a discourse segment is a dynamic property: while ARCs may be at issue at a certain point in a discourse (e.g., when they are the most recent unit), they may no longer be at-issue at a later point.

(11) a. Todd's boat, which is a Santa Cruz 27, is fun to race. $\rightarrow$ preference for indirect rejection

b. We had fun racing on Todd's boat, which is a Santa Cruz 27. $\rightarrow$ direct rejection more likely

Ultimately, the extent to which this is due to a discourse- or a processing-based constraint is unclear. It may simply be that once the processing of a sentence-medial discourse segment like an appositive is complete and a speaker has returned to the main clause content, all else equal, comprehenders will choose to relate incoming content with more recent content. However, the ARC may remain accessible so long as there is an appropriate cue to backtrack to a previous segment (like the presence of *actually* in (10), which signals an upcoming correction). Nevertheless, we will see that there are interesting online processing differences between sentence-medial and sentence-final ARCs. I return to a discussion of this in §3.1.2.

Taken together, these diagnostics complicate the empirical picture. While ARC content appears to be independent in some sense, it's not the case that it is always inaccessible. Taking these and other diagnostics into account, some researchers have moved away from semantic explanations in favor of discourse-based ones. For example, AnderBois et al. (2010) appeal to conversational dynamics, suggesting that while at-issue content introduces a proposal to update the common ground, not-at-issue content automatically updates the common ground. Simons, Tonhauser, Beaver, and Roberts (2010) propose a different type of pragmatic explanation: that a proposition is at-issue if it is relevant, in the sense of Roberts (1996/2012), to the current Question Under Discussion (QuD). Under this approach, an appositive can be at-issue if it directly addresses the QuD. For example, the ARC in (13) is not-at-issue with respect to the QuD in (12a) because it is irrelevant but perhaps offers an explanation behind the main assertion of the sentence, whereas (13) is at-issue with respect to the complex QuD in (12b) because it offers an explanation that responds directly to the conjoined question (*Why are the sailors sailing back into the harbor? Because they see the coming storm*). Therefore, while appositives may contribute discourse asides in the sense that they are irrelevant in the formal sense, they may of course be related to a discourse in the sense that they offer elaborations to the main point of an utterance.

(12) a. QuD: What are the sailors doing?

b. QuD': What are the sailors doing and why?

(13) The sailors, who see the coming storm, are sailing back into the harbor.

Others suggest that appositives are asides in the sense that they contribute discourse-subordinate relations, in the sense of Asher and Lascarides (2003). Indeed, in the example above, the ARC is linked to the main content through an Explanation relation. However, this is not a ubiquitous or defining property of appositives. Koev (2013) notes that ARCs may be linked to their hosts via Narration relations, as in (14). Here, each unit sequentially addresses the subQuDs: *What happened at time t1?*, *What happened at time t2?*, *What happened at time t3?* Therefore, the discourse units are taken to be linked by coordinating relations.

(14) Nick stuck out his tongue at Jamie. Jamie, who (then) hit him, left in a huff.

One open issue concerns whether it is ultimately easier to establish a coordinating or subordinating relationship between main clauses and appositives. At present, we don't know enough about the range of possible discourse contexts that appositives appear in to answer this question (see Loock, 2007, for more detail). For the time being, I simply note that ARCs are not always discourse subordinate.

3.1.1.2 Syntactic Properties

So far, I've discussed various semantic and pragmatic properties of appositives that have led to the claim that they are independent at one or both of those levels of representation. Arguments involving the syntactic independence of appositives are similarly fraught. Some researchers suggest that syntactically, ARCs are independent root clauses (McCawley, 1998), whereas others maintain a syntactic relationship between the main clause and the ARC (Jackendoff et al., 1977). All accounts must allow for some type of relationship between the RC head and the RC, although some maintain that this is purely an anaphoric relationship (Ott, 2016).

There are some compelling arguments for the syntactic independence of ARCs. Take, for instance, the following examples adapted from Burton-Roberts (1999). The constituents of idioms can be separated by an RRC boundary (15a) and still retain their idiomatic meaning, but cannot be separated by an ARC boundary (15b). The relative pronoun may be absent in RRCs but not ARCs (16). ARCs, but not RRCs, may serve as fragment responses anaphoric to a previous cross-speaker utterance in a discourse (17). Finally, a variety of syntactic constituents are permitted as antecedents to ARCs, whereas RRCs can only modify NPs (see examples in Burton-Roberts, 1999).

(15) a. The cat that Leela let out of the bag concerned your demotion.

b. #The cat, which Leela let out of the bag, concerned your demotion.

(16) a. The article I'm working on needs to be short.

b. *The article, I'm working on, needs to be short.

(17) A: My publications will include the article in the Scientific American.

B: {Which, *That} you've not even begun to write yet.

At the same time, though, other examples suggest that ARCs are not completely syntactically independent. As we saw earlier, they may feature cross-boundary verb-and noun-phrase ellipsis (which I assume, in part, requires access to a syntactic representation), like in (5). In German and Turkish, they often feature case-matching between the ARC head and its internal contents (Potts, 2005). This suggests that appositives cannot be completely syntactically independent, as certain syntactic relations are permitted to span their boundaries. The following section (§3.1.2) also presents processing evidence against their syntactic independence. These facts evidencing some type of intact syntactic relationship are difficult to reconcile with the claim that ARCs are entirely independent root clauses.

3.1.1.3 Prosodic Properties

One property of ARCs that is not in dispute, and is particularly relevant to the hypothesis proposed in §2.3.3, is their prosodic independence (in English). Researchers generally agree that appositives are phrased in separate intonational phrases, which feature some combination of the following prosodic features: pauses surrounding their left and right boundaries, pre-boundary lengthening, boundary tones, and in some cases, differences in intonational contour, pitch compression, or changes in speech rate of appositive content. While it is unlikely that there is one unique parenthetical intonation assigned to ARCs (c.f. Potts' (2005) "comma intonation"), corpus and production studies have found constancy in the presence of pre- and post-ARC boundaries. In a corpus study, Dehé (2009) found that ARCs are phrased in distinct intonational phrases the vast majority of the time. This was diagnosed using a set of ARC-internal and external criteria. Internal criteria included presence of a complete intonational contour, pitch declination within the ARC domain, and pitch reset immediately following ARC boundaries. Pitch reset at boundaries was also used to diagnose pitch compression of the ARC relative to the main clause. External criteria diagnosing cues to boundaries between intonational domains included pre- and post-boundary pauses (only those clearly aligned with structural boundaries, not hesitations or disfluencies), phrasal and boundary tones in their surrounding environment, and domain-final lengthening.

91% of Dehé's data evidenced clear intonational boundaries on either edge of the appositive unit, as evidenced by the presence of the prosodic features mentioned above. In the remaining 9% of cases, ARC content was less clearly separated, in the sense that the right boundary was always clearly demarcated, but the relative pronoun following the left boundary was phrased along with preceding main clause material. In an example like (18), for instance, the relative pronoun contained an L% boundary tone on the relative pronoun *which*, the absence of a pause between the left syntactic main-RC boundary, but a presence of a 185 ms pause on average following the relative pronoun.

(18) ($_\iota$I was programming in Pascal which) ($_\iota$really isn't very exciting) ($_\iota$I'm afraid)
H*L-L% H*L-L%

Although this provides evidence of a syntax-prosody misalignment in a small number of cases, it should be noted that all instances of ARCs identified in the corpus involved prosodic separation of ARC content. I thus take Dehé's results as evidence that ARCs are indeed reliably prosodically separated in naturally occurring speech. Dehé's analysis does not discuss the intonational contours of appositives in depth, beyond noting that there is a tendency to copy the preceding phrasal and boundary tones in the host clause onto the right edge of the ARC. This argument is supported by one example, with intonational contours as annotated in (18).

Although Dehé's investigation of naturally occurring speech validated the prosodic separation of ARCs, laboratory-based production studies have been less successful in finding evidence of clear prosodic demarcation of ARCs. In a small production study with 6 speakers, Hirschberg and Avesani (1997) find a greater degree of variability in the prosodic phrasing of appositives. They provide speakers with relative clauses, not

	# of isolated RCs	% of isolated RCs
Non-restrictive context	9/18	50%
Restrictive context	4/18	22%

Table 3.1: Rates of prosodic isolation for ambiguous RCs in Hirschberg and Avesani (1997).

marked as appositives with commas, but instead biased towards RRC or ARC interpretations via a preceding restrictive or non-restrictive context. Their results, in Table 3.1, suggest a weaker tendency to prosodically isolate ARC content; however, they still find that ARCs are more likely to be isolated than RRCs. Some caution should be exercised in over-interpreting their results, as the small number of tokens for each structure (n = 18) makes it difficult to draw definitive conclusions.

Watson and Gibson (2004) extended Hirschberg and Avesani experiment (N = 11) using a partner communication task. They, too, used ambiguous RC structures with no orthographic cues to ARC boundaries that were biased towards ARC or RRC readings by the context. They also investigated whether prosodic separation varied by whether the RC was subject- vs. object-modifying, as in (20).

(19) a. *Non-Restrictive S1:* A group of film critics praised a director and a producer.

b. *Restrictive S1:* A group of film critics praised a director at a banquet and another director at a film premiere.

(20) a. *Subject-Modifying S2:* The director **who the critics praised at a banquet** insulted an actor from an action movie during an interview.

b. *Object-Modifying S2:* An actor from an action movie insulted the director **who the critics praised at a banquet** during an interview.

The results (Table 3.2) suggest that pre-RC boundaries were relatively unlikely, regardless of preceding context. In fact, sentence-position seemed more important than context in determining the likelihood of prosodic isolation: object-modifying RCs

	Subject-modifying	Object-modifying
Non-restrictive context	27.7%	49.1%
Restrictive context	10.0%	38.6%

Table 3.2: Pre-RC boundaries in Watson and Gibson (2004), Expt. 3.

were more likely to be accompanied by a pre-RC boundary than subject-modifying RCs. The authors suggest that this may be due to the fact that RCs appearing late in a clause are more likely to be interpreted as non-restrictive.

The fact that these studies fail to obtain clear evidence for prosodic demarcation of appositives may be driven by the fact that they used ambiguous relative clause structures, without the canonical orthographic marking of an ARC. Thus, the lack of prosodic separation may simply reflect the fact that participants ultimately were unlikely to interpret RCs without clear cues to prosodic boundaries as being non-restrictive. To my ear, a subject-modifying RRC following the context in (19a) sounds perfectly acceptable, especially if the RC head bears contrastive accent, like in (21), to indicate the director as opposed to the producer within the set of individuals praised by the film critics. Alternatively, participants may have interpreted the head of the RC in S2 as a different director from the one introduced in the context sentence: the director they praised at the banquet, in contrast to the director they praised elsewhere.

(21) A group of film critics praised a director and a producer. The diRECtor that the film critics praised insulted an actor.

Watson and Gibson (2004) do not report information about the focal structure of participants' productions (i.e., which constituents bore focal pitch accents), so it is not possible to empirically validate these possibilities. But, Astruc-Aguilera and Nolan (2007), another production study that did mark ARCs with canonical punctuation, find results compatible with Dehé's (2009) claims: in addition to a strong tendency to prosodically isolate ARC content, they report additional evidence for the contour copying (matching intonation on the pre-parenthetical region and the right boundary of the parenthetical itself) noted by Dehé.

In conclusion, appositives are clearly prosodically separated from main clause content in naturally occurring productions of unambiguous structures. The studies on ARC processing that this dissertation builds on were conducted using reading methodologies. Therefore, we require an understanding of how the overt prosody of ARCs relates to their implicit prosody. It is notable that in written text, the only surface difference between ARCs and RRCs is sometimes the presence of commas. In the productions studies discussed here, the presence of a comma in text was sufficient evidence for producers to prosodically demarcate appositive content in speech. I therefore reason based on the available evidence that ARCs receive a distinct implicit prosodic representation as well. That is, the presence of orthographic cues to appositive boundaries in written text are crucial to their prosodic representation, and perhaps also affect their information status as secondary discourse segments (see §3.1.1.4).

Empirically, the experimental studies discussed in §3.1.2 along with the novel results reported in this dissertation support this assumption: evidence suggests that prosodic cues to ARC boundaries are effectively utilized by the processor. The Context-Sensitive Encoding Hypothesis assumes that the implicit prosodic structure of ARCs is one crucial source of evidence that can be used to drive the segmentation of these units during comprehension and the formation of encoding contexts in memory. But ultimately, further production and listening studies on ARCs must be conducted in order to fully assess the role that (implicit) prosody plays in their processing (see §6).

3.1.1.4 Discourse-Prosody Interactions

Work on intonation establishes a meaningful interaction between discourse and prosodic factors in determining sentence meaning. Before concluding this section, then, I turn to a proposal by Westera (2019) concerning the relationship between the intonational and discourse properties of appositives and other parenthetical structures. Westera (2019) starts from the observation that ARCs are a type of construction that can bear a rise-fall-rise (RFR) intonational contour, as in (22), then proposes that this signals a particular relation to the Question Under Discussion (QuD) structure of a discourse.

(22) John, who is a vegetarian, envies Fred.
L*H-H% L*HL-H% H*L-L%

Specifically, Westera proposes that an utterance containing an appositive or parenthetical may pursue multiple QuDs: a primary topic of discourse (addressed by the main clause) and a secondary topic (addressed by the parenthetical content). Therefore, a speaker may employ a particular intonational strategy, like rise-fall-rise, in order to signal to their interlocutor that they are pursuing a secondary QuD, which does not directly relate to the primary topic. Under this view, guided by Westera's (2017) Theory of Intonational Compliance Marking (ICM), the use of RFR indicates non-compliance with typical conversational maxims (Grice, 1975) relative to the current (primary) QuD. Specifically, the intonational components (i.e., the phrasal and boundary tones) of the RFR contour are assumed to signal particular pragmatic meanings, as in (23).

(23) **Assumption 1**: (Westera, 2019)

a. L% / H%: the speaker {takes, doesn't take} the utterance (up to this boundary) to comply with the maxims with respect to the main QUD.

b. -L / -H: the speaker {takes, doesn't take} the utterance (up to the first subsequent boundary tone) to comply with the maxims with respect to some focus-congruent QUD.

Under (23), the intonational contour on (22) conveys that: (i) the utterance up until the left edge of the parenthetical does not (yet) fully address the main QuD but does address a secondary QuD, as indicated by the presence of the L-. This theory makes two predictions. First, it predicts that appositives should not be permitted to bear H- phrase accents or L% boundary tones, because these components are taken to signal non-compliance with a secondary QuD or compliance with the main QuD, respectively. Secondly, it predicts that the left edge of ARCs is not permitted to bear an L% boundary tone (unless sentence final), because the main QuD should be unresolved at this point[15].

The appeal of Westera's theory is twofold: (i) it assumes that in a coherent discourse, the relation between main and ARC content is not random, but rather that the interlocutors of a conversation seek to construct discourse relations between the units, and (ii) it further assumes that a speaker is likely to opt for an intonational strategy that will clearly signal a discourse strategy to their interlocutor (i.e., akin to a meta-conversational move indicating something along the lines of *Wait a moment – I'm addressing a secondary question, but I'll get back to the main point momentarily*). Nevertheless, it is unclear whether the ICM theory of ARCs allows for other types of intonational contours (other than RFR) and other types of discourse relations (e.g., NARRATION). If the theory maintains that appositive intonation should always map to a RFR contour, this is likely too restrictive. Example (18) from Dehé (2009) suggests that the left and right boundaries of ARCs may be marked with L% tones, which would suggest that that both units address the main QuD under Westera's proposal. In the case that this example is treated like a sentence-final ARC though, an L% boundary tone at the left edge may be permitted. In unpublished pilot data examining production of sentence-medial nominal appositives (joint work with Jack Duff), we

[15] Although, it is unclear what this theory predicts about continuative appositives, where both the main clause and ARC address a main QuD *What happened?* in narrative contexts with subquestions like: *What happened at t1?, What happened at t2?, What happened at t3?*

also find that the right edge of ARCs may reliably feature L% boundary tones. Thus, it is likely that different intonational contours are permitted and may serve to signal different pragmatic functions (e.g., elaboration, continuation, identification, etc.). In line with this view, Auran and Loock (2011) have established that in British English, different discourse functions of appositive relative clauses (continuative, subjectivity, and relevance ARCs) relate to different degrees of prosodic separation, where continuative ARCs are less discontinuous than other types; this further suggests that not all appositives pattern the same in their discourse-prosody mapping. Nevertheless, an extension of Westera's proposal may prove to be capable of describing a wider variety of contours and contexts. However, there is likely significant individual variability in the particular appositive intonation that a speaker may choose to employ during conversation, or in their inner voice during reading; therefore, any theory that posits a one-to-one mapping between particular intonational contours and pragmatic functions will likely prove too restrictive.

At a higher level, Westera's theory highlights that oftentimes, changes in prosodic properties go hand in hand with changes in information structure. In the case of appositives, prosodic boundaries whether implicit or overt may naturally lead to ARCs being interpreted as discourse asides (see Duff et al., 2023, for a related discussion). Thus, it may be difficult (perhaps, impossible) to attribute the independence of ARCs to their prosodic separation alone. This does not mean that their prosodic separation does not play an important role in their ability to be treated as independent during sentence comprehension; indeed, the current results support this conclusion on empirical and conceptual grounds (see §3.1.2). Whether implicit prosodic boundaries (in the absence of a preceding context) are likely to also systematically induce a particular discourse effect on ARCs is an open question. We've seen that there is at least some degree of variance in the discourse status of these constructions. It's possible that the default preference is to expect a subordinate relationship between ARC and main clause content. What we don't know is how early comprehenders make such discourse commitments, or what evidence they use to do so.

3.1.2 Processing Profile

Work on the processing of appositives has revealed two related observations that are often attributed to a common source. Following Duff et al. (2023), I term the first observation the *discounting effect*: that in offline acceptability judgments, appositive relative clauses (ARCs) receive higher ratings of acceptability or naturalness than restrictive relative clauses (RRCs). As such, ARCs have been argued to contribute less to judgments of perceived syntactic complexity (Dillon et al., 2014, 2018; Duff et al., 2023; Kroll & Wagers, 2019). I term the second observation the *bypassing effect*: that during online dependency resolution, intervening ARCs appear to contribute less to retrieval interference than RRCs do, leading to the conclusion that the content of ARCs is bypassed in some sense by memory retrieval operations. Although previous studies have implicitly assumed that discounting and bypassing arise from a common source, they also vary widely in claims about the nature of the memory mechanisms involved. In this chapter, I examine the relationship between discounting and bypassing. I ultimately argue that while both can be derived from a common temporal context mechanism that enforces partitions in the memory representation, discounting and bypassing arise from distinct sources. Specifically, I suggest that the dynamics of reinstating previous encoding contexts during comprehension (per Reinstantiation) drive bypassing, whereas contextual dissimilarity during encoding (per Context-Sensitive Encoding) contributes to discounting.

Previous work reveals disagreement on another dimension, namely which level of linguistic independence is best equipped to capture the processing profile of apposi-

	SHORT	LONG
RRC	That evil man who was on the cruise tried to intimidate the waitress.	That evil man who was on the cruise Mary took to the Pacific Islands tried to intimidate the waitress.
ARC	That evil man, the one who was on the cruise, tried to intimidate the waitress.	That evil man, the one who was on the cruise Mary took to the Pacific Islands, tried to intimidate the waitress.

Table 3.3: Example item set from Dillon et al. (2014), Experiment 3.

tives. This section reviews the mechanistic and linguistic arguments put forward in previous work, ultimately concluding that discourse-based explanations of discounting and bypassing fall short in their explanatory potential. Other studies argue against appeals to syntactic independence (Dillon et al., 2018), but offer support for the role of prosodic independence (Kroll & Wagers, 2019). I conclude by revisiting the premises of Reinstantiation: that the prosodic independence of appositives in tandem with top-down syntactic expectations may interact with a Context-Sensitive Encoding mechanism to derive bypassing effects.

3.1.2.1 Discounting

Let's return to the observation that ARCs and nominal appositives are discounted relative to the integrated content of restrictive relative clauses (RRCs), like in Table 3.3 (Dillon et al., 2014). That is, while the RRC-SHORT and ARC-SHORT conditions are rated as equally acceptable, ARC-LONG is significantly more acceptable than ARC-SHORT. This has led to the conclusion that while additional length generally negatively affects acceptability, this complexity penalty is ameliorated for appositives.

Dillon et al. (2014) attribute the discounting effect to the fact that while appositives contribute independent speech acts to their containing sentences, RRCs do not.

Structure	Length: SHORT, (LONG)
DD	Evan said that the cruise (Mary took to the Pacific Islands) departed three hours behind schedule.
ID	Evan said, "the cruise (Mary took to the Pacific Islands) departed three hours behind schedule."

Table 3.4: Example item set from Duff et al. (2023), Experiment 2.
Key: DD = direct discourse; ID = indirect discourse.

Following up on this idea, Duff et al. (2023) attempt to search for an analogous pragmatic discounting effect by comparing acceptability of two types of speech reports: direct discourse, which always contributes an independent speech act, and indirect discourse, which does not contribute an independent speech act and is syntactically integrated with the matrix clause. They tested the Speech Act Discounting Hypothesis: that once complete, speech acts are discounted in later parsing and decision-making. Their design crossed Structure (DD, ID) with Length (SHORT, LONG), as in Table 3.4. Like in the ARC case, Speech Act Discounting should predict a reduced complexity penalty for the LONG-ID condition.

Duff et al. fail to find evidence for the interaction predicted by Speech Act Discounting. They conclude that it is not the independent speech act status of appositives that contributes to their discounting. In a second attempt to explore evidence for pragmatic discounting, they entertain the idea that the status of appositives as independent discourse segments, in the sense of Asher and Lascarides (2003), allows for their discounting. They test the Discourse Unit Discounting Hypothesis: that completed discourse segments are discounted in later parsing and decision-making processes. Their design utilized the contrast between two constructions: non-restrictive *because*-clauses, which contribute independent discourse segments, and restrictive *when*-clauses, which are not discourse-independent, as in Table 3.5. Here, too, Duff et al. do not observe the interaction predicted under Discourse Unit Discounting, once

Structure	Length: SHORT, (LONG)
when	Evan often complains to the travel agent when storms delay the cruises (Mary takes to the Pacific Islands).
because	Evan often complains to the travel agent because storms delay the cruises (Mary takes to the Pacific Islands).

Table 3.5: Example item set from Duff et al. (2023), Experiment 5.

more failing to support the idea that the discourse independence of appositives drives their discounting.

Other studies have attributed discounting to the independence of ARCs in terms of their Question Under Discussion (QuD) structure (S. Kim & Xiang, 2022). S. Kim and Xiang suggest that ARCs address secondary QuDs because they are discourse-subordinate, and their processing behavior can be attributed to the fact that memory retrieval only targets discourse segments associated with the current or active QuD; therefore, they postulate a +SUBORDINATE feature in memory, motivated by the fact that ARCs are argued to be discourse-subordinate to the host clause they attach to. It should be noted though that although they find evidence of ARC bypassing (see a discussion of these experiments in the following section), their sentences are presented out of context, and there is no empirical evidence to validate the types of QuDs participants may be reconstructing for their items. Additionally, as discussed in §3.1.1.1, ARCs need not be discourse-subordinate. Recall that within an SDRT framework, continuative ARCs (Holler, 2008; Jasinskaja, 2016; Koev, 2013) stand in a coordinating (non-subordinate) relationship with host material.

In an unpublished manuscript, Kroll and Wagers (2019) directly investigate the role of QuDs in ARC processing using complex questions in combination with preceding discourse contexts, which force appositive content to address the main QuD of an utterance, as in (24). They hypothesized that if ARCs are discounted because of a preference to associate them with secondary QuDs, ARCs that address the main QuD should not display discounting.

(24) QuD: Where is the bear standing and what is it wearing?
The bear (who is standing on the ball the trainer rolled across the room) is wearing a hat.

Kroll and Wagers' results show that modulating the at-issue status of appositive content does not cause length-based complexity penalties to emerge for appositives. That is, they still find that long ARCs are discounted, even when they directly address the QuD. Thus, their discounting behavior cannot be explained by a propensity to address a secondary QuD.

Although many researchers have pursued the intuition that ARC discounting stems from a pragmatic source, systematic investigations have revealed little to no evidence in support of this position. Along three dimensions of pragmatic independence (speech act status, discourse segmentation, and association with a main vs. secondary QuD), previous studies have failed to obtain evidence that appositive discounting is a product of semantic or pragmatic independence. In each of these cases, ARCs have the ability to *sometimes* constitute independent speech acts, or secondary discourse units, but they do not always do so.

It could be argued that the status of ARCs as discourse-subordinate units (or "backgrounded" units in some more general sense) is perhaps less flexible, but the soundness of this claim is somewhat uncertain. Recall Redeker's finding (§2.1.3.2) that there is a general dispreference to treat segments as parenthetical in the absence of explicit evidence to do so, and furthermore, that this decision is not made on the basis of structural or lexical cues alone. In other words, the discourse context in combination with what Redeker terms "transitional cues" were jointly used to determine the relevant discourse relationship. It may be that ARCs are different from the parentheticals Redeker used. Their syntactic and prosodic structures may always make them more prone to being interpreted as discourse-subordinate. However, the timecourse of interpretation is relevant here. Some researchers argue that sentence-medial ARCs are most likely to be interpreted as discourse-subordinate, whereas sentence-final ARCs are more likely to receive a variable interpretation, but one where at-issue status is more easily achieved (Jasinskaja, 2016; Syrett & Koev, 2015). Under these approaches, the discourse status of ARCs is a property that changes over time, but not in a strictly incremental fashion; they are marked as subordinate when they are no longer the final segment. I return to this point in the context of work on ARC bypassing.

Only one set of studies has attempted to explicitly investigate the role of implicit prosody in discounting. Kroll and Wagers (2019) find support for the idea that discounting is due to the prosodic independence of appositives. They compared the processing of long ARCs vs. RRCs where each segment was visually separated via a line break, a method which was previously been used by Swets et al. (2007) to encourage implicit prosodic breaks in a relative clause attachment reading study. They used stimuli as in (25).

(25) The bear
that is standing on the ball the trainer rolled across the room
is wearing a hat.

In this experiment, they found that the complexity difference between ARCs and RRCs collapsed; long, visually segmented RRCs were perceived to be no more complex than long, visually segmented RRCs. A follow-up study on sentence-final visually demarcated RRCs led to the same conclusion. On this basis, Kroll and Wagers (2019) advance a mechanistic explanation for discounting: they suggest that the prosodic segmentation of a sentence influences the allocation of working memory resources. If processing resources are allocated to a sentence in a segment-by-segment manner, the amount of syntactic content actively maintained in working memory at any given time should be reduced for ARCs compared to their RRC counterparts. This offers an explanation of discounting based on prosodic structure that is not dependent on the syntactic position of appositives, in contrast the fact that Dillon et al. (2014) The explanation I put forward for discounting, namely Context-Sensitive Encoding, will mirror the spirit of Kroll and Wagers's proposal closely, but it pins discounting on a slightly different mechanistic source: greater distinctiveness in contextual representation.

Although the results of the line break experiment point to a prosodic explanation, another two experiments reported by Kroll and Wagers (2019) that attempted to induce prosodic breaks at only the left or right edge of RRCs with adverbials (26) did not result in RRC discounting.

(26) a. Left boundary: The bear is wearing a beret that, predictably, is a light blue color the French trainer picked out.

b. Right boundary: The bear who is standing on the ball the trainer rolled across the room is, predictably, wearing a hat.

Taken together, Kroll and Wagers' prosodic independence studies suggest that prosodic cues to both the left and right boundaries of an appositive play a necessary role in discounting. Dillon et al. (2014) rule out a prosodic explanation for a related reason. They surmise that there should be a preference for a prosodic break following a long subject-modifying restrictive relative clause, which would lead to a decrement in acceptability for these structures. But because final, object-modifying RRCs also evidence discounting, they conclude that a prosodic explanation is insufficient. However, Kroll and Wagers's (2019) account offers a reasonable explanation as to why prosodic segmentation should lead to a benefit for both subject- and object-modifying ARCs in acceptability judgments, only when both boundaries are explicitly marked. If the ability to demarcate a unit as prosodically independent during encoding affects the memory load associated with any given segment, this should lead to a greater percept of acceptability for more segmented sentences, regardless of the syntactic position of an appositive. By the same token, if subject-modifying ARCs doubly benefit from reduced processing load due to prosodic segmentation and from reinstatement of main clause material, we should expect subject ARCs to be more acceptable than object ARCs, which partition their containing sentences into fewer segments and do not involve reinstatement. In line with this view, Dillon et al. (2014) find that subject ARCs are rated as more acceptable than object ARCs. I will return to a more in depth discussion of these points in §3.1.3.

Of course, direct discourse, one of the structures used by Duff et al. (2023) in their investigation of discounting, is also prosodically distinct from its matrix clause. Investigations of its implicit prosody show that readers are sensitive to the presence of quotations in text. In auditory perceptual simulation studies, where readers "mentally simulate characteristics of voices attributed to a particular speaker or a character depicted in the text" (Zhou & Christianson, 2016), readers subconsciously modulate reading speed of quotations according to imagined properties of the speaker (e.g., their speech rate and accent) (Yao & Scheepers, 2011; Zhou & Christianson, 2016). In addition, functional magnetic resonance imaging studies reveal higher brain activity in speech-related areas of the auditory cortex for direct discourse compared to indirect discourse, supporting the idea that a reader's experience of the inner voice is more vivid during reading of quotes (Yao, Belin, & Scheepers, 2011).

Recall from the discussion of Duff et al. (2023) above, though, that their two offline studies on direct discourse did not find evidence of discounting. Under an account of discounting that relies on prosodic segmentation, this remains somewhat puzzling. We offer a potential explanation for this. First, it should be noted that some work suggests that speakers' prosody in naturally occurring speech does not reliably or differentially demarcate the boundaries of direct vs. indirect speech reports (Klewitz &

Couper-Kuhlen, 1999). That being said, in reading, there should still be other cues to boundaries (e.g., the visual cues provided by quotation marks, and possibly a shift in the nature of the implicit prosodic representation). These features should in principle make direct discourse a prime candidate for a construction that reliably shifts contextual representations, but segmentation and contextual distinctiveness are not the only factors that should be expected to influence acceptability judgments in these cases. Though direct discourse is represented more vividly, and is correspondingly remembered better in verbatim recall than indirect discourse (Eerland, Engelen, & Zwaan, 2013), this may not straightforwardly relate to an increase in acceptability. If readers expend more effort encoding direct discourse, this may lead to a decrease in acceptability, which cancels out the general benefit of segmentation in acceptability judgments, while still contributing to a segmentation benefit in memory. We tentatively conclude then that prosodic segmentation offers a promising explanation for discounting of ARCs, and that a combination of other factors may influence acceptability ratings for direct speech reports.

Finally, previous studies fail to support an account of discounting that relates to syntactic independence of ARCs alone. In an offline judgment study, Dillon et al. (2018) test the Subordinate Clause Hypothesis, that the ARC creates a syntactic constituent with the host noun it attaches to. In NP of PP RC attachment ambiguities with restrictive relative clause constructions as in Table 3.6, *with*-PPs generally lead to a higher proportion of low attachment responses than *of*-PPs do (see Frazier and Clifton (1996) and Dillon et al. (2018) for a thematic domain-related explanation). If ARCs always attach high because they are independent clauses, Dillon et al. (2018) hypothesized that they should not display sensitivity to the *of* vs. *with* distinction. In line with the Subordinate Clause Hypothesis, however, the results suggested a greater propensity to attach low in *with*-PP constructions regardless of RC type, suggesting

Structure	Sentence
RRC	Penny ignored the child {of, with} the patient that had an annoying voice.
ARC	Penny ignored the child {of, with} the patient, who had an annoying voice.

Table 3.6: Example item set from Dillon et al. (2018), Experiment 1.

that ARCs are treated as syntactically subordinate to the host clause. The following section, which reviews studies on bypassing of ARCs online, discusses additional processing evidence for (limited) syntactic interactivity between the main and ARC clauses.

In sum, the studies summarized here provide little support for the idea that ARCs are discounted on account of their discourse or syntactic independence alone, but offer some promising evidence in favor of a role for prosodic independence.

3.1.2.2 Bypassing

In studies on online comprehension, ARC content appears to be bypassed: it contributes less to retrieval interference than RRC content during dependency resolution. In particular, ARCs have been shown to interfere less during the resolution of filler-gap dependencies (Dillon et al., 2017), agreement processing (S. Kim & Xiang, 2022; McInnerney & Atkinson, 2020), and anaphora (S. Kim & Xiang, 2023) than RRC content does.

In their investigation of filler-gap processing, Dillon et al. (2017) found that retrieval and integration of a filler with a gap site across an intervening ARC was less costly (i.e., resulted in shorter go-past times in eye-tracking) than across an intervening RRC. Their design, which crosses Dependency (+WH, -WH) with Structure (RRC, ARC), is given in Table 3.7.

It is standardly assumed that in filler-gap dependencies, the filler (*who*) is actively

	+WH	-WH
RRC	The butcher asked <u>who</u> the lady who bought Italian ham had invited _ to dinner tonight.	The butcher asked if the lady who bought Italian ham had invited anyone to dinner tonight.
ARC	The butcher asked <u>who</u> the lady, who bought Italian ham, had invited _ to dinner tonight.	The butcher asked if the lady, who bought Italian ham, had invited anyone to dinner tonight.

Table 3.7: Example item set from Dillon et al. (2017), Experiment 3.

maintained in working memory until the parser reaches a syntactically appropriate gap site that is not otherwise filled (Frazier, Clifton, & Randall, 1983). Dillon et al. (2017) assume that at this point, the filler is retrieved from memory and integrated into the gap site[16]. Their proposal is that retrieving a filler across an intervening ARC is less costly because the contents of the ARC are rapidly lost, or compressed, in memory once its processing is complete. This compression of structure is further assumed to be triggered by the subordinate discourse status of the ARC; specifically, the authors assume that its contents are less likely to be necessary for later processing operations. This aligns with the idea that the encoding format of previous sentence content may change at certain points during processing, such that the syntactic representation is no longer actively maintained (c.f. Parker & Phillips, 2016). We term this proposal *Compression*, and offer the definition in (27).

(27) **Compression**: After the processing of a sentence-medial ARC, its syntactic representation is suppressed in memory due to the independent discourse status of ARCs.

Crucially, Dillon et al. (2017) do not observe any difference in reading behavior between ARCs and RRCs until past the right boundary of the appositive. At the left

[16]If a complete representation of the filler is actively maintained, it is not clear that any retrieval operation must take place at the gap site. Perhaps Dillon et al. (2017) assume that retrieval is necessary because active gap-filling is suspended at the left edge of an appositive boundary. This detail is underspecified in their proposal, but I set this aside for the purpose of the present discussion.

boundary of an ARC, one might predict a "switch cost" for shifting attention from the main clause to the appositive while holding a filler in memory. Such an effect does not emerge at the RC-initial region (*who bought*). In fact, their data reveals a trend in the opposite direction, that the +WH-RRC condition yields longer first-pass reading times at this region; however, this effect does not reach significance. At the region containing the gap site (*(anyone) to dinner*), Dillon et al. find that filler-gap integration following an ARC is less costly than when following an RRC. This leads them to suggest that Compression is a process that applies once the processing of an ARC unit is complete. In line with this suggestion, in an eye-tracking study on wrap-up effects (the tendency for readers to dwell longer on sentence-medial positions marked by punctuation) (Hirotani et al., 2006) found a wrap-up cost on the right boundary of ARCs, but not on the left boundary. Hirotani et al. (2006) attribute these results to the prosodic properties of appositives alone (that they are phrased in separate ιPs), but recall that Dehé (2009) and others find that both the left and right edges of ARCs are reliably prosodically separated in speech. Thus, it is unlikely that Hirotani et al.'s reading time results are a product of prosodic separation alone.

Crucially, a Compression-based account of ARCs predicts worse syntactic memory for appositive content and more difficulty with accessing ARC-internal contents during dependency resolution. Dependencies spanning host-appositive boundaries are certainly attested (AnderBois, Brasoveanu, & Henderson, 2015), but no work to date assesses their ease of processing online. To my knowledge, only one study (Dillon et al., 2018) has experimentally investigated access to ARC-internal content using acceptability as a proxy for antecedent accessibility, but this study does not confirm the predictions of Compression. I return to a discussion of this shortly. Note also that work on prosodic segmentation might predict the opposite pattern of results: that the prosodic segmentation of ARCs results in better memory for their content, per

Exp.	Attr?	Structural Configuration
Exp. 1	✗	*The waitress, who _ sat near **the girls**, **were** unhappy.
Exp. 2	✔	***The musicians**, who the reviewer **praise** _ so highly...
Exp. 3	✔	*Alicia met **the musicians**, who the reviewer **praise** _ so highly...

Table 3.8: Structural configurations from S. Kim and Xiang (2022), Experiments 1-3. 'Attr?' indicates whether the experiment found an agreement attraction effect.

Context-Sensitive Encoding (CSE), and facilitated access to the contents of appositives under certain circumstances, per Context-Sensitive Retrieval (CSR). §3.1.3 discusses these predictions in detail. Thus, Compression and CSE/CSR offer two (non-mutually exclusive) sets of predictions for Experiments 1-3 reported in this chapter.

Other studies have similarly observed an asymmetry in the directionality of interference in sentences containing ARCs: that prior to the right boundary of an ARC, main content may interact with ARC content, but past the right boundary, ARC content does not interfere with main content. For instance, agreement attraction effects obtain in the Main $\rightarrow$ ARC direction, but not the ARC $\rightarrow$ Main direction (S. Kim & Xiang, 2022; McInnerney & Atkinson, 2020). S. Kim and Xiang (2022) establish this using the contrast in Table 3.8, where a potential agreement attraction configuration either spanned an ARC-RRC boundary or an RRC-ARC boundary.

Agreement attraction refers to an effect where a number-matching but structurally illicit intervener in complex subjects like *the key that opened* ***the cabinets****...* in (28) leads to facilitation of reading times for ungrammatical subject-verb agreement, as in the continuation ***were*** *rusty* (Wagers et al., 2009).

(28) *The key that opened the **the cabinets** unsurprisingly **were** rusty.

In S. Kim and Xiang's (2022) Experiment 1, where a number-matching distractor was contained within the ARC and the critical verb (*praises*) subsequently occurred in the main clause, no agreement attraction effect obtained. In contrast, Experiments

2-3 showed that a preceding distractor in the initial segment of the main clause was accessible to the RC verb, and resulted in agreement attraction. Overall, this pattern is consistent with the left vs. right edge asymmetry observed for ARCs (c.f. Dillon et al., 2017). However, I note a potential issue with comparing the structural configurations in Experiments 1-3. In Experiment 1, the head noun *waitress* is integrated into a subject gap site within the relative clause; thus, the potential agreement attraction effect at *were* can only be due to retrieval interference at this region, which the results do not support. In Experiments 2-3, however, the head of the object relative clause (*musicians*) must be maintained until the RC-internal gap site, which occurs in the position immediately after the RC verb. Thus, the agreement attraction effects observed in these experiments may be driven by interference from the active maintenance of the distractor at the time of parsing the verb. Therefore, the particular contrast in Table 3.8 may not be well-equipped to establish the directionality of interaction between RRCs vs. ARCs. Nevertheless, the results of their Experiment 1 offer additional evidence to support bypassing once ARC processing is complete.

Moreover, the results reported by Hirotani et al. (2006) and Dillon et al. (2017) suggest a processing difference at the left vs. right boundaries of ARCs. In all of these cases, there is a syntactic relationship in the Main $\rightarrow$ direction (between the RC head and the gap site) prior to the right edge of the ARC, but past this right boundary, there is no ARC $\rightarrow$ Main relationship, only a main clause RC head-verb relationship. This raises a question about what drives contextual encodings in sentences containing ARCs: the syntactic relationships between each of the clauses, or their prosodic separation. The experiments in Chapter §4 will ultimately offer evidence for the latter.

In another self-paced reading experiment, S. Kim and Xiang (2023) find that ARC content is bypassed in ambiguous anaphoric contexts. Their design (in Table 3.9) varies the degree of interference from a potential ARC-internal antecedent by manipulating

Structure	Position	Sentence
RRC	MED	**The violinists** who admired **the singers** invited **their mentors** to the party.
ARC	MED	**The violinists**, who admired **the singers**, invited **their mentors** to the party.
RRC	FIN	**The singers** admired **the violinists** who invited **their mentors** to the party.
ARC	FIN	**The singers** admired **the violinists**, who invited **their mentors** to the party.

Table 3.9: Example item set from S. Kim and Xiang (2023; 2024).

its plurality, then measures reaction times at the ambiguous pronoun region (*their mentors*). Because this pronoun may refer to either the main or ARC antecedent, they assume that the retrieval probe itself does not privilege a particular discourse segment (S. Kim & Xiang, 2024). At the pronoun, they found faster reaction times in the MED-ARC condition than in the MED-RRC condition, further supporting the idea that ARC content interferes less. Additionally, they found that sentence-final ARCs and RRCs exhibited no difference in reaction times to the same decision. This reinforces the claim that there is a difference in processing operations following left vs. right ARC boundaries.

The authors suggest that the penalty for medial RRCs is due to encoding interference across similar discourse segments, both marked with a +MAIN feature in memory. In contrast, the medial ARC condition, marked with a +SUBORD feature in memory, yields less interference. This contrast is exemplified in (29)-(30). The authors attribute this to encoding (not retrieval) interference due to the fact that the interpretation of the pronoun is ultimately ambiguous, and so the retrieval probe is not specified for discourse-related cues. In a follow-up visual world study, S. Kim and Xiang (2024) further advance this argument by noting that competition between the NPs arises within the RC region itself, not in the main clause. That is, images of violinists and singers received equal looks during processing of the RC in their RRC, whereas the difference between looks to either antecedent was greater in their ARC condition. When additional length was added to the end of the RC, this competition effect did not spill over into the main verb or pronoun region.

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

+MAIN	+MAIN	+MAIN

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

+MAIN	+SUBORDINATE	+MAIN

(S. Kim & Xiang, 2023, 2024)

I take no issue with the suggestion that similar NPs in RRC structures may engender greater competition within the RC than in ARC structures. In spirit, this basic premise accords well with the Context-Sensitive Encoding Hypothesis proposed here: that prosodic boundaries trigger shifts in encoding context, and that the complexity penalty for RRCs can be attributed to greater contextual similarity between items during encoding. However, CSE does not assume item-level encoding interference between main and ARC content, e.g., through overwriting of item-level features (Nairne, 1990). Evidence for encoding interference of this sort has standardly come from temporally proximal or overlapping, similar items (Gordon, Hendrick, & Johnson, 2001; Rich, 2024; Smith, Franck, & Tabor, 2021; Villata, Tabor, & Franck, 2018), not between subjects and objects in subject RC constructions. As such, the particular hypothesis that I invoke in §3.1.3 (Reinstantiation) will attribute the ARC bypassing effect to reduced retrieval interference, not encoding interference. I argue also that S. Kim and Xiang's results do not clearly point to encoding interference as the source of bypassing, which by definition, is an effect observed at the main clause retrieval site. If their results related to the bypassing standardly observed for ARCs, we should expect to see the difference in competition persist into the main clause as well, at the main verb[17]

[17]Because a subject must be retrieved at this point, the main clause NP should already be pre-

and pronoun regions. Thus, their results suggest that the segments of sentences containing ARCs feature separation during encoding, but they do not support the view that the bypassing effect itself is due to reduced encoding interference.

I therefore suggest that any explanation of bypassing that relies on memory retrieval operations selecting for particular discourse segments must frame interference effects in terms of retrieval, not encoding. But, a discourse status-based explanation runs into two other problems. First, if the retrieval probe does not specify cues to particular discourse segments, then there must be some process of downweighting the activation of retrieval candidates in discourse-subordinate units during the search process. A potential timecourse is sketched in (31). Such a process might predict a larger cost for ARCs, due to the need to suppress activation of ARC-internal candidates, but this is the opposite of the observed pattern of results. It would also require that we rework our understanding of cue-based retrieval more generally, as standard conceptions assume only stage ③. A more plausible explanation is that discourse status is used to filter feature-matching candidates at a later stage, not during the retrieval process itself.

(31) Proposed timecourse for MED-ARC condition from S. Kim and Xiang (2024)

Retrieval cues on **their mentors**: {+NP, +PL}

① **Access stage**, where the search for matching candidates returns:

{+NP, +PL, +MAIN}, {+NP, +PL, +SUBORD}

② **Suppression stage**: {+NP, +PL, +SUBORD} is downweighted

③ **Selection stage**: {+NP, +PL, +MAIN} is selected due to greater activation

The second problem relates to the timecourse of marking an ARC as discourse-subordinate. The earliest evidence of differential processing for ARCs appears on their right boundary (Hirotani et al., 2006). It could be argued that this is the lo-

activated prior to the pronoun.

cus of encoding their subordinate discourse status. However, coherence relations are determined according to the relationship between entire discourse segments. At the point of reaching the right boundary of the ARC, the comprehender has encountered very little information about the main clause (only its subject). This is ostensibly not enough information to determine whether an ARC will contribute subordinating or coordinating information relative to the main clause. Determining secondary status with respect to the main QuD runs into the same issue: upon encountering only the subject of the main clause, relevance of the ARC with respect to the main QuD (in the absence of context) cannot definitively be established.

As suggested in §3.1.2.1, it could be that there is a default preference to always expect that an ARC will contribute a subordinate/secondary unit. However, we do not have specific empirical evidence to support the position that this commitment takes place in advance of encountering sufficient main clause content. At least some studies suggest that establishing discourse relations is delayed until processing of the relevant segments is complete (Duff, 2023; Millis & Just, 1994). The behavior of ARCs would have to diverge from these better-studied cases in order for immediate commitment to a subordinate relationship to take hold. This is not an impossibility, if different constructions involve establishing coherence relations in different ways.

The different accounts of bypassing summarized here rely on different underlying mechanisms. Dillon et al.'s (2017) Compression account suggests that syntactic memory for ARC content is rapidly lost at its right boundary. A modified version of S. Kim and Xiang's (2024) account might suggest that the preference for establishing dependencies across main discourse units manifests as the presence of a +Main feature on the retrieval probe[18]. The following section entertains these possibilities in

[18] This seems unlikely, given work on anaphoric processing that divides the process into two stages: retrieval and resolution (Garrod & Sanford, 1982). This suggests that the memory retrieval mechanism should target all feature-matching items, but filter activated candidates based on other discourse factors at a later stage. Nevertheless, I consider this possibility for the sake of completeness.

more detail. Regardless of which type of mechanism is adopted, these accounts make the prediction that retrieval of ARC-internal content should be more difficult than that of RRC-internal content once processing of the ARC is complete.

An acceptability judgment reported by Dillon et al. (2018) comments on the accessibility of ARC-internal discourse anaphoric antecedents. Their item sets were like in Table 3.10, where *one*-anaphora within a sentence-final nominal appositive referred to a host clause antecedent or to an ARC-internal antecedent. Their design crossed the structure of the object RC (RRC, ARC) with the referent of the *one*-anaphor (SUBJ, OBJ). Recall that Dillon et al. (2017) suggest that main clause content remains more accessible over an intervening ARC than over an intervening RRC. Therefore, if ARC-internal antecedents are more difficult to access than RRC-internal ones, and if acceptability ratings are reflective of ease of discourse anaphora in the current study, Dillon et al. (2018) should have observed a crossover interaction such that the SUBJ-ARC condition received higher ratings than the SUBJ-RRC condition due to less interference in the ARC case, and the OBJ-ARC condition was rated lower than the OBJ-RRC condition due to compression of ARC content. Their results revealed only a main effect of antecedent (an OBJ preference over SUBJ) and a main effect of Structure (an RRC preference over ARC) but no significant interaction. Although their OBJ-ARC condition was rated numerically higher than their SUBJ-ARC condition, they nevertheless take the results to "…fit more naturally with the view that the availability of the not-at-issue material alone is diminished after it has been understood" (Dillon et al., 2018, p. 11), perhaps due to the fact that the OBJ-ARC condition was rated numerically lower than the OBJ-RRC condition.

However, this conclusion is not entirely consistent with their results, which suggest that SUBJ vs. OBJ accessibility was not meaningfully different for ARCs and RRCs. Furthermore, an unpublished study by Sharf (2024) reports that appositive-internal

	SUBJ	OBJ
RRC	**That girl** was at the party that was held in the big auditorium, **the one** who you wanted to date.	That girl was at the party that was held in **the big auditorium**, **the one** that has a capacity of 10,000.
ARC	**That girl** was at the party, which was held in the big auditorium, **the one** who you wanted to date.	That girl was at the party, which was held in **the big auditorium**, **the one** that has a capacity of 10,000.

Table 3.10: Example item set from Dillon et al. (2018), Experiment 2.

antecedents are successfully primed by the presence of a co-referential pronoun, providing tentative support for the conclusion that ARC content remains accessible. This calls into question the validity of approaches that involve Compression, although it remains possible that there is a general preference to search for main clause content. I argue, however, that the bypassing effect can be derived without resorting to explanations that assume ARC content is less accessible; the following section explains how.

3.1.3 Context-Sensitive Memory for Appositives

Previous accounts of the processing of ARCs make (i) commitments about the encoding of different segments and (ii) predictions about the status of appositive content in memory and its resulting accessibility, but do not empirically verify (i) or (ii). This section details how the Context-Sensitive Encoding (CSE) Hypothesis, repeated in (32), can account for potential mechanisms driving the discounting and bypassing behavior of appositives. In addition, the Context-Sensitive Retrieval (CSR) Hypothesis considers whether contextual interference has an effect during online comprehension. In particular, it investigates whether cue-based retrieval of an item from within an appositive is facilitated relative to retrieval of an item from an RRC due to reduced contextual interference. I return to a discussion of CSR shortly.

(32) **Context-Sensitive Encoding** (CSE) **Hypothesis**: Prosodic boundaries shift the *temporal context vector* (Howard & Kahana, 2002) during encoding, such that prosodic phrases partition syntactic content in memory into distinct *encoding contexts* (EC), and the content within an EC shares similar contextual features. These ECs are built up incrementally and may be reactivated on a by-context basis at a later stage, using positional cues. Segmented sentences therefore incur less processing load than non-segmented ones, due to decreased contextual interference during encoding and retrieval.

Recall from §2.3.2 that the TCM assumes that items are bound to a slowly-evolving context vector during encoding. Following Wagers (2008), I've proposed that at prosodic boundaries, the context vector shifts such that items preceding a shift are more contextually similar to one another than items following a shift. In principle, this has several potential consequences. First, memory for sentences with greater segmentation should display a benefit relative to non-segmented sentences, because the less syntactic material associated with a given context, the less contextual interference the encoding or retrieval of that context should incur. This accords well with the discounting effect observed for ARCs, as well as the general segmentation benefit for prosodic phrases reviewed in §2.1.2.3.

During the online processing of sentences with ARCs, CSE hypothetically allows for the remaining mechanisms in Table 3.11: Reinstantiation, Backgrounding[19], Compression, and CSR. Each of these make different predictions about (i) the reasons for discounting and bypassing and (ii) the accessibility of content internal to an appositive. The *ARC Access* column contains predictions regarding point (ii); ✔ indicates facilitated access to ARC-internal content (relative to RRC-internal content), and ✗ in-

[19]As stated in §1, the *Backgrounding* mechanism is meant to stand in as a theory-neutral cover for any account that assumes ARCs are interpreted as peripheral, in some sense, to the main point of the discourse. This includes discourse-subordinate approaches under SDRT, relation to a secondary QuD under QuD theory, or other more general notions of backgroundedness.

Mechanism	ARC Access	Definition
CSE	✔	See definition in (32).
Reinstantiation (to be revised)	=	The contents of an earlier encoding context (= main clause) can be reinstated using positional cues in order to add additional material to a previously incomplete segment (c.f. Sederberg et al., 2011), such that the initial and final segments of the main clause in sentences containing ARCs are encoded as contextually similar because they have a syntactic relationship.
Backgrounding	✗	The contents of the current (= ARC) encoding context can optionally be reactivated and marked with a context-level feature like +BKGRD (given the appropriate discourse context).
Compression	✗	The contents of the current (= ARC) encoding context may be reactivated and then compressed in memory such that its syntactic representation is degraded (given the appropriate discourse context).
CSR	✔	See definition in (35).

Table 3.11: Potential mechanisms underlying ARC processing.
ARC Access Key: ✔: access to ARC-internal content is facilitated; ✗: access to ARC-internal content is inhibited; =: equal access to ARC- vs. RRC-internal content.

dicates inhibited access to ARC-internal content. I assume that Reinstantiation, Backgrounding, Compression, and Context-Sensitive Retrieval each depend on a CSE-like mechanism, which allows the parser to demarcate contextual domains by incrementally shifting the context at significant structural boundaries. Each of these approaches is discussed below. Part of the challenge with disentangling these possibilities is that they are not necessarily mutually exclusive mechanisms. Nevertheless, the evidence presented throughout the remainder of the dissertation will serve to rule out some but not others.

CSE accounts for discounting, as segmentation should reduce memory load associated with encoding each segment and thus reduce the perception of complexity for both sentence-medial and sentence-final ARCs (c.f. Dillon et al., 2014). On its own, however, it cannot account for bypassing; a pure effect of prosodic boundaries cannot explain why access to main content appears to be privileged in sentences with ARCs. Reinstantiation differs from a pure effect of prosodic boundaries by assuming that at certain points during incremental processing, an earlier context may be reactivated for the purpose of adding additional content (c.f., Grosz & Sidner, 1986). This amounts to the difference between the initial and final segments of the main clause being treated as part of the same encoding context, as in (34), as opposed to being treated as distinct encoding contexts, as in (33).

(33) CSE

1 2

The cat, who loved cheese, was taking a nap.

1 : *cat,* → context shift

2 : *cheese,* → context shift

(34) CSE + Reinstantiation

1 2 3

The cat, who loved cheese, was taking a nap.

1 : *cat,* → context shift

2 : *cheese,* → context shift, reinstate initial context

3 : ... → continue parsing, with initial context relatively more active

Reinstantiation assumes the following. Once processing of the ARC is complete (at its right boundary), top-down syntactic expectation for a continuation of the incomplete main clause leads to reinstatement of the contextual state associated with the initial encoding context, retrieved via a positional cue to that context. This accounts for the wrap-up cost observed only at the right edge of ARCs (Hirotani et al., 2006)[20] and the fact that subsequent access to main clause material is facilitated (Dillon et al., 2017; S. Kim & Xiang, 2022, 2023). Recall the claims from §3.1.2.2 that not all ARCs are discourse-subordinate, and that the marking of a subordinate discourse segment may not occur immediately. From this point of view, Reinstantiation offers good explanatory potential for an account of the bypassing effect that is not reliant on the discourse status of ARCs. If the presence of a sentence-medial, ARC-final comma cues pre-activation of the context associated with the main clause in advance, main content should then be relatively more active immediately prior to processing subsequent main clause dependencies. In this case, main clause content in sentences with ARCs will always become relatively more active at the right edge of the ARC regardless of whether the ARC is subordinate. Such an approach is ultimately better-equipped to explain why ARCs are so reliably bypassed.

Reinstantiation thus makes several predictions: (i) that main content should be relatively more accessible past the right ARC boundary, (ii) that access to ARC-internal content should be no more difficult than access to RRC-internal content given sufficiently informative item-level cues, and (iii) that main content should ultimately be more accessible due to contextual reinstatement. The bypassing effect validates point (i), and the results of Experiments 2-3 of this chapter are consistent with prediction (ii). Chapter 4 will establish that other constructions involving sentence-medial prosodic phrases also display evidence of bypassing, ultimately validating prediction (iii). Beyond the conceptual arguments provided here and in the previous section, the current chapter does not empirically test the predictions of Reinstantiation. Thus, I will return to a more thorough discussion of this process later.

Backgrounding (c.f., S. Kim & Xiang, 2022, 2024) and Compression (c.f., Dillon et

[20] Although reactivation and/or compression of the current context's contents under Backgrounding or Compression could account for this cost as well.

al., 2017) assume that some process applies at the right edge of ARCs that renders their contents relatively less accessible later on due to their subordinate discourse status. Backgrounding and Compression both predict that once the parser has passed the right edge of an ARC, retrieving ARC-internal content should be more difficult than retrieving RRC-internal content, albeit for different reasons. As mentioned above, I assume that both processes must rely on a mechanism like CSE: in order to ensure that a process operates over only the relevant linguistic domain, the mechanism must have access to some cue to the left edge of the ARC. Under CSE, the initial context shift at the left boundary provides a positional marker that ensures such operations proceed over only the current encoding context at the point of completing the appositive. While Compression predicts that ARC-internal content should be degraded after its processing is complete, Backgrounding simply predicts a preference to retrieve material from the main clause. The experiments reported in this chapter show that ARC-internal content is better represented than RRC content and is equally as accessible as RRC-internal content. Thus, they rule out Backgrounding and Compression.

Any of the three mechanisms discussed so far (Reinstantiation, Compression, Backgrounding) can account for discounting of ARCs, because CSE underlies each of them. CSE and Reinstantiation also predict better retention of syntactic memory for ARC content compared to RRC content, because ARCs should be encoded as contextually distinct from other sentence content. Backgrounding does not make specific claims about the strength of ARC content in memory, whereas Compression assumes that the representation of ARC content should be degraded.

Finally, the Context-Sensitive Retrieval Hypothesis (35) considers the effect of encoding context shifts on cue-based retrieval during the resolution of item-to-item dependencies.

(35) **Context-Sensitive Retrieval** (CSR): the cue-based retrieval mechanism can use the context vector as a cue, because retrieval of an item reinstates that item's encoding context. Therefore, retrieval of an item from a larger encoding context (i.e., one that contains more lexical/syntactic content) incurs greater contextual interference than that of a smaller encoding context.

Context-Sensitive Retrieval predicts that retrieval of an item from a sentence containing an ARC (in the absence of cue overload[21]) should be facilitated relative to that of an RRC-containing sentence, because contextual interference should be reduced for prosodically segmented structures, if similarity of contextual features is a potential source of interference following item-level retrieval. Experiments 2-3 use noun phrase ellipsis (NPE) in structures like (37)-(38) to test this, but do not find substantial evidence in support of CSR.

NPE is a type of ellipsis in which the head of certain DPs (specifically, numerals, quantifiers, and possessives) licenses the omission of a following noun. As such, NPE may be licensed by a numeral, like in (36). We chose to use NPE for this experiment is because it targets a small constituent (a noun phrase) that can be contained within the relative clause structures we manipulated.

(36) Imala has two cats, and Georgie has four __.

CSR predicts that the resolution of NPE in sentences with sentence-medial (37) and sentence-final (38) ARCs should be facilitated relative to sentences with non-prosodically segmented RRCs. This contrasts directly with the predictions of Backgrounding and Compression, under which accessing ARC-internal content should be more difficult.

[21]In instances of cue overload (e.g., potential antecedents in both the main clause and the appositive that match the features specified by the retrieval probe), we observe bypassing. Therefore, in order to test for CSR, we had to set up a scenario where a retrieval probe in the main clause uniquely targeted a dependent inside the appositive.

(37) 1 2 3

The author, | who wrote **two novels**, | met the hack that wrote **forty** _

1 : *author,* → context shift

2 : *novels,* → context shift

3 : *forty* _ → retrieve *novels*, reinstate its encoding context

(38) 1 2

The author that wrote **two novels** met the hack, | who wrote **forty** _

1 : *hack,* → context shift

2 : *forty* _ → retrieve *novels*, reinstate its encoding context

For the sake of reducing the possible hypotheses on the table, Experiment 1 contrasts the predictions of CSE with Compression by investigating the retention of syntactic memory for ARC content, ultimately finding evidence more consistent with CSE. Experiments 2-3 then turn to investigating the accessibility of ARC-internal content, contrasting the predictions of Compression and Backgrounding on the one hand with CSR on the other. The results provide only weak evidence for CSR, and are inconsistent with Compression. As mentioned previously, these mechanisms need not be mutually exclusive; the results leave room for the possibility that a process like Backgrounding may play a role alongside CSR. The discussion of the results further elaborates on this point.

3.2 Experiment 1

Experiment 1, a recognition memory study, sought to determine whether syntactic memory for ARCs is better (per Context-Sensitive Encoding; CSE) or worse (per Compression) than that of RRCs. The data were subjected to a Signal Detection Theoretic analysis (Hautus, Macmillan, & Creelman, 2021) in order to compare sensitivity (d_a)

to ARC content with sensitivity to RRC content. The predictions are straightforward. Under CSE, we predicted that the structure of ARCs should be better remembered than that of RRCs. Under Compression, we predicted the opposite, namely that ARCs should be worse remembered than RRCs. Finally, we hypothesized that it is possible that both mechanisms are concurrently active and may conspire to cancel one another out; in this case, we expected to find no difference in sensitivity between structures.

3.2.1 Method

3.2.1.1 Participants

54 participants were recruited via Prolific to participate in the experiment. Participants were located within the USA, their first language was English, they had completed at least a high school diploma or equivalent, and were at least 18 years old; they were also restricted to users with a 90% or higher approval rate on Prolific with a minimum of 20 submissions. Each experimental session took approximately 45-60 minutes, and participants were compensated $12/hour for participation. Participants with 75% or less comprehension question accuracy were excluded (N = 6). The 48 remaining participants were included in the analysis.

3.2.1.2 Materials

48 item sets crossed two factors: the RC Structure (rrc, arc) with Match (match, mismatch), which manipulated whether the syntactic structure between an exposure and target sentence matched or mismatched. Match and mismatch conditions used ditransitive RC constructions with dative/double object alternations, in order to ensure that the syntactic structure of the RCs varied while holding meaning constant. For half of the items, exposure sentences contained dative constructions within the relative clause, and the structure of the relative clause in the test sentence either matched

	MATCH	MISMATCH
RRC	The father *that cooked the kids a meal* after the orchestra performance was grateful for instant noodles.	The father *that cooked a meal for the kids* after the orchestra performance was grateful for instant noodles.
ARC	The father, *who cooked the kids a meal after the orchestra performance*, was grateful for instant noodles.	The father, *who cooked a meal for the kids after the orchestra performance*, was grateful for instant noodles.

Table 3.12: Example item set from Experiment 1.

(also contained a dative construction) or mismatched (contained a double object construction). For the other half of the items, the exposure RCs contained double object constructions. A sample item set is given in Table 3.12, where the MATCH column corresponds to the exposure sentence participants were presented with. Participants then saw a MATCH or MISMATCH target sentence, depending on the condition.

3.2.1.3 Procedure

All experiments reported in the dissertation were coded in Ibex (Drummond, 2010) and presented using PCIbex (Zehr & Schwarz, 2018).

In the exposure phase of each trial, participants read visually chunked sentences on a screen at their own pace using a cumulative, chunk-by-chunk self-paced reading procedure. Each prosodic phrase was chunked together such that the comma was not the sole cue to prosodic phrasing; however, each chunk remained on the screen as the next one was revealed. An example of this presentation method is given in (39).

(39) _________ ________________ __________

The father, ________________ __________

The father, who cooked a meal, __________

The father, who cooked a meal, was grateful.

Following the exposure, participants performed a distractor task in which they were asked to solve a randomly generated arithmetic problem. These problems involved addition or subtraction of two numbers between 0 and 50, where the answer was always a positive value. After typing the answer in a box and submitting it, the target sentence was presented in its entirety. This sentence either matched the exposure exactly, or the syntactic structure of the medial clause varied. Prosodic phrasing (cued by comma presence) was always identical across exposure and target presentations. Participants were instructed to provide a 'Same' or 'Different' decision on the target sentence, prompted by the question *Was this the same sentence you saw before?* After responding to this question, participants were asked to rate their confidence in their answer on a scale of 1-3 (where 1 was 'not at all confident', 2 was 'somewhat confident', and 3 was 'very confident') on the next screen.

Experimental items were presented across 4 Latin-squared lists along with 108 fillers. 48 of these fillers involved detecting changes to subject vs. object-relative clauses within a medial phrase of prosodically isolated and integrated coordinate structures. Another 60 fillers of varied structures, both segmented and non-segmented, involved detecting dative/double object alternations and changes to the temporal order of constituents. Fillers varied the location of changes such that participants could not learn to only attend to the medial region (the region of interest in our critical sentences).

3.2.2 Results

3.2.2.1 Signal Detection Theory Analysis

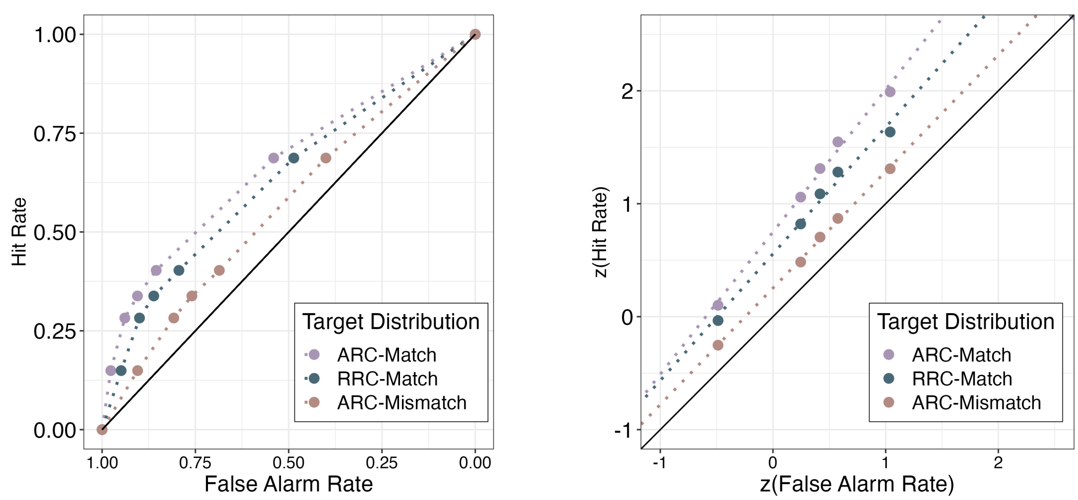
The analysis utilized unequal variance Signal Detection Theory (SDT) (Hautus et al., 2021) to determine how sensitive participants were to detecting changes to the syntactic structure of ARCs vs. RRCs. SDT has classically been applied to recognition memory experiments, as it serves as a psychological model of decision making that allows one to derive a measure of sensitivity termed d' (i.e., how well participants are able to distinguish between target (MATCH) and lure (MISMATCH) conditions) independent of the likelihood of responding 'Same' vs. 'Different' (i.e., a participant's *response bias*). SDT additionally accounts for the fact that there are two ways for participants to contribute a "correct" response. Participants may correctly respond *Same* to a sentence that they have encountered before (a 'Hit'), or they may correctly respond *Different* to a sentence they have not encountered before (a 'Correct Rejection'). It is assumed that each of these response types is drawn from a different underlying distribution: a target distribution, consisting of responses to the original exposure stimuli, and a lure distribution, consisting of responses to novel distractors not previously encountered by participants.

SDT offers a measure of the overall *sensitivity* to a particular condition, termed d_{a}, independent of the level of response bias a participant may display for responding 'Same' vs. 'Different'. d_{a} is measured in terms of root mean squared standard deviation units (as opposed to d', which is measured in simple standard deviation units and assumes equal variance between the H and FA-generating distributions). Both d' and d_{a} are interpreted as a measure of the perceptual distance between the target and lure distributions, independent of response bias, but d_{a} additionally allows for unequal variance in the underlying distributions. Given a particular perceptual distance between the H and FA distributions (the distance between the mean of the H-distribution and the mean of the FA-distribution), a participant may choose set a *decision criterion* at any point along that interval. As the decision criterion moves, the perceptual distance between the distributions remains constant, but the propensity for a participant to respond *Same* vs. *Different* shifts. Using a set of empirically derived sensitivity and response bias levels, one can create a *receiver operating charac-*

teristic (ROC) curve, which shows the different criterion locations a participant may adopt (derived via a combination of Same/Different decisions and confidence ratings on each of those decisions) given some constant measure of sensitivity or overall performance on the task. Plotting multiple ROCs allows one to visualize differences in sensitivity across conditions.

3.2.2.2 SDT Results

Overall accuracy for experimental and filler items was reasonably high (85%), and accuracy for the experimental items alone was 95%. Results are reported in Table 3.13 and plotted in Figure 3.1.



(a) Experiment 1 empirical ROCs. **(b)** Experiment 1 zROCs.

Figure 3.1: Empirical ROCs derived from raw Hit and False Alarm rates (Fig. 4.8a) and zROCs derived from z-transformed H and FA rates (Fig. 4.8b) for Experiment 1.

Sensitivity, in terms of d_a, for each condition was calculated by combining the Same/Different judgements and confidence ratings into a six-point scale, ranging from *Very Confident-Same* to *Very Confident-Different*. This resulted in the empirical ROCs plotted in Figure 4.8a, where each condition was scaled against the RRC-MISMATCH condition. Corresponding zROCs are plotted in Figure 4.8b. Area-under-the-curve

(AUC) values for ARC vs. RRC conditions were calculated using the `pROC` package in R (Dillon & Wagers, 2019; Robin et al., 2011). Statistical comparison of AUC values was performed with `pROC`, using 2000 bootstrap replicates.

	d_a	AUC	2.5%	97.5%
ARC	0.84	0.67	0.64	0.7
RRC	0.64	0.63	0.6	0.66
	D_{boot} = 1.76		p = 0.08	

Table 3.13: `pROC` model results for Experiment 1.

Results reveal that sensitivity to ARC structures (d_a = 0.84) is higher than for RRC structures (d_a = 0.64), suggesting that the structure of ARCs is better remembered than RRCs. This is ultimately more consistent with the predictions of CSE. We fail to find support for Compression, which predicts the opposite pattern of results; that is, ARCs should be worse remembered under this hypothesis. However, `pROC` model results indicated that the difference in AUC values was only marginally significant ($p = 0.08$). This is discussed further below.

3.2.3 Discussion

The results are straightforward: we find a pattern of results inconsistent with Compression, and generally more supportive of CSE. Sensitivity to ARCs was indeed higher than that of RRCs, which does not align with the idea that the syntactic contents of ARCs are rapidly lost in memory (c.f. Dillon et al., 2017). This suggests that in offline memory representations, the benefit bestowed by the prosodic segmentation of ARCs outweighs any detrimental effect their subordinate status may have on memory. Furthermore, the results support the foundational premise of CSE that prosodic boundaries influence separation of main vs. ARC content during encoding.

One could potentially argue that the fact that ARCs were only marginally better remembered may lend partial support for Compression. That is, the benefit of prosodic segmentation per CSE and the degradation of ARC content per Compression may be co-active and cancel one another out, but the prosodic segmentation benefit slightly outweighs the effect of Compression. If Compression is at play, we should expect that online access to ARC-internal content is inhibited. We suggest based on the results of Experiment 2 that this is unlikely.

Another potential concern involves the chunk-by-chunk reading method we employed in the presentation of the exposure sentence. Per Kroll and Wagers's (2019) results, visual segmentation may have had an independent effect on the prosodic phrasing readers imposed on the experimental sentences. However, in the context of this experiment, we argue that the presentation did not drive our results for two reasons. First, note that the presentation of chunks was cumulative. We opted to present exposures in chunks to encourage participants to thoroughly read the sentence, and to adopt the desired first-pass prosody on our critical sentences. As each new chunk was revealed, the previous ones remained on the screen. The reading of the exposure was also untimed, so participants were free to re-read the full sentence as many times as they desired. After the last chunk was revealed, the full sentence was presented with only commas as cues to the prosodic phrasing and structure of the sentence. Thus, there was ample time to override the initial presentation method with the preferred sentence-final implicit prosody.

Secondly, we conducted an analysis of 48 filler item sets with the design in Table 3.14, which crossed the Phrasing of a medial list conjunct (ISO, INT), whether that item was prosodically isolated or integrated with previous content, with Match (MATCH, MISMATCH), whether the structure of that item was a subject or object relative clause. The boundaries of cumulatively presented chunks aligned with comma positions for each of the exposure sentences. This resulted in the same difference in number of

	MATCH	MISMATCH
ISO	Kenji was introduced both to the chef with many awards, *and to the baker that loved the chef*, but not to the apprentice with the diligent work-ethic at the gala over the weekend.	Kenji was introduced both to the chef with many awards, *and to the chef that the baker loved*, but not to the apprentice with the diligent work-ethic at the gala over the weekend.
INT	Kenji was introduced both to the chef with many awards *and to the baker that loved the chef*, but not to the apprentice with the diligent work-ethic at the gala over the weekend.	Kenji was introduced both to the chef with many awards *and to the chef that the baker loved*, but not to the apprentice with the diligent work-ethic at the gala over the weekend.

Table 3.14: Example item set from subset of Experiment 1 fillers (item n = 48).

chunks for ISO vs. INT conditions (3 vs. 2) as was present in the critical ARC vs. RRC exposures.

Results revealed no difference in sensitivity between ISO and INT conditions, despite the fact that their chunking visually mirrored that of the critical items. The contrast between this analysis and the analysis of our critical items suggests that our visual presentation method was not the driving force behind our results. Furthermore, it may suggest that orthographic cues to implicit phonological phrase boundaries (40) are not sufficient to drive contextual delineation, whereas the intonational phrase boundaries of ARCs are. This departs from the assumptions of a mechanism like Visibility (see §2.2.1), where φPs are taken to be sufficient to influence attachment decisions.

(40) ($_{\varphi}$Kenji was introduced both to the chef with many awards)
($_{\varphi}$and to the baker that loved the chef)
($_{\varphi}$but not to the apprentice with the diligent work-ethic)

3.3 Experiment 2

Experiment 2 aimed to determine how the presence of an ARC affects access to main vs. ARC content during cue-based retrieval. The experiment assessed this by measuring response latencies to sentence regions that required the resolution of noun phrase ellipsis (NPE) in a reading study utilizing the Maze task, a word-by-word reading methodology (see §3.3.1.3).

The experimental design contrasted the predictions of Compression/Backgrounding with Context-Sensitive Retrieval. The predictions are as follows. Compression assumes that once the processing of an ARC is complete, its contents are rendered less accessible. Backgrounding relies on a preference to resolve dependencies with main-clause antecedents over ARC-internal antecedents. Thus, the predictions of Compression and Backgrounding were the same in the present experiment. Both hypotheses predict more difficulty accessing ARC-internal content than RRC-internal content (i.e., slower response latencies), specifically for sentence-medial ARCs. When an ARC is sentence-final, Compression and Backgrounding assume that accessing earlier main clause content should be equally easy.

Recall that Context-Sensitive Retrieval predicts the clearest effect in the case of a strong retrieval cue to the antecedent; thus the design of Experiment 2 utilized lexical and structural parallelism leading up to the antecedent and ellipsis site as an unambiguous cue to the antecedent, so as to best observe the effect of boundaries while minimizing other sources of item-level retrieval interference. Under these circumstances, CSR predicts facilitated access to any segment in a sentence containing an ARC. That is, we should expect to observe faster response latencies to ARC-internal antecedents in sentence-medial ARC configurations, and faster response latencies to main clause antecedents in sentences with sentence-final ARCs, compared to respective RRC-only baseline sentences. This is because CSR assumes that segmentation in the absence of cue overload serves to reduce the amount of contextual interference that accompanies the retrieval of any given item, compared to the amount of potential contextual interference in a non-segmented structural counterpart.

These predictions are reiterated shortly in the context of the experimental items (§3.3.1.2).

3.3.1 Method

3.3.1.1 Participants

90 Prolific participants took part in the experiment and were paid $12/hr. Participants were located within the USA, their first language was English, they had completed at least a high school diploma or equivalent, and were at least 18 years old; they were also restricted to users with a 90% or higher approval rate on Prolific with a minimum of 20 submissions. Any participant who failed more than 50% of Maze trials was excluded. 20% of participants were excluded on this basis, leaving 72 participants in the analysis. Though comprehension questions were included in order to ensure that participants had incentive to pay attention to the sentences, participants were not excluded on the basis of comprehension question accuracy. Because the Maze task forces highly incremental processing, we assumed that merely completing the sentence without failing out of the trial was taken as sufficient evidence that participants were attending to the stimuli (see §3.3.1.3 for more detail).

3.3.1.2 Materials

36 item sets (Table 3.15) featured nominal ellipsis that either spanned a restrictive relative clause boundary, as in the CONTROL condition, or appositive relative clause boundaries, as in the ARC-1 and ARC-2 conditions.

Structure	Sentence **critical region** | spillover region
CONTROL	The struggling author that published two novels resented the successful hack that published forty <~~novels~~> **over** the past three decades.
ARC-1	The struggling author, who published two novels, resented the successful hack that published forty <~~novels~~> **over** the past three decades.
ARC-2	The struggling author that published two novels resented the successful hack, who published forty <~~novels~~> **over** the past three decades.

Table 3.15: Example item set from Experiment 2.

Condition	Backgrounding, Compression	CSR
ARC-1	✗ (ARC > RRC)	✔ (ARC < RRC)
ARC-2	ARC = RRC	✔ (ARC < RRC)

Table 3.16: Predicted reaction times under Compression and CSR for Experiment 2. ✔/✗ indicate facilitation/inhibition, respectively, relative to the CONTROL condition.

In ARC-1, a post-ARC ellipsis site in the main clause required retrieval of an antecedent from within the appositive. In ARC-2, an ARC-internal nominal ellipsis site required retrieving an antecedent from the host. The CONTROL condition contained two restrictive relative clauses, the second of which contained the nominal ellipsis site and the first of which contained the antecedent. The number of words (and amount of structure) contained in the antecedent region varied slightly across experimental items. The antecedent in Table 3.15, for instance, is 2 words (*two novels*), whereas in other cases, it was longer (*two pieces of candy*). Across items, this region always contained between 2-4 words. Reaction times were measured in a Maze task at the critical region (the word immediately following the ellipsis site, bolded above) and the spillover region (the two following words, underlined above). <~~Elided content~~> is indicated in Table 3.15, but was not presented to participants during the experiment. The predictions for each condition under Compression and CSR are presented in Table 3.16.

Experimental items were counterbalanced across 3 Latin-squared lists and presented along with 136 fillers. Fillers included list constructions containing a mix of ellipsis and anaphora, other relative clause constructions (e.g., NP of NP RC), and subordinate clause structures containing noun phrase ellipsis.

3.3.1.3 Procedure

Experiment 2 utilized the A-Maze (Boyce, Futrell, & Levy, 2020; Forster, Guerrera, & Elliot, 2009), a word-by-word reading task that forces participants to choose between a grammatical and a high-surprisal continuation. An example trial is depicted in Figure 3.2. At the beginning of a trial, participants are given a choice between the first word and a string of x-es (x-x-x). For every subsequent word, they must choose between two continuations via a button press. In Figure 3.2, for example, *The struggling...* constitutes a grammatical continuation, whereas *The accomplish...* does not. In this particular implementation, if a participant selected the wrong continuation, the trial ended and immediately proceeded to a post-sentence comprehension question, or continued on to the next trial.

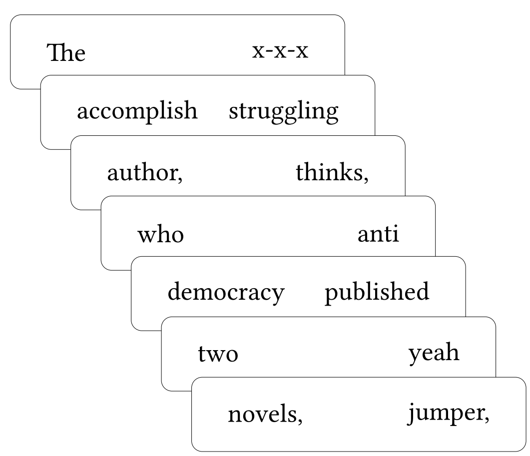


Figure 3.2: Example Maze trial from Experiment 1.

Response latencies collected from this task are taken to index linguistic processing

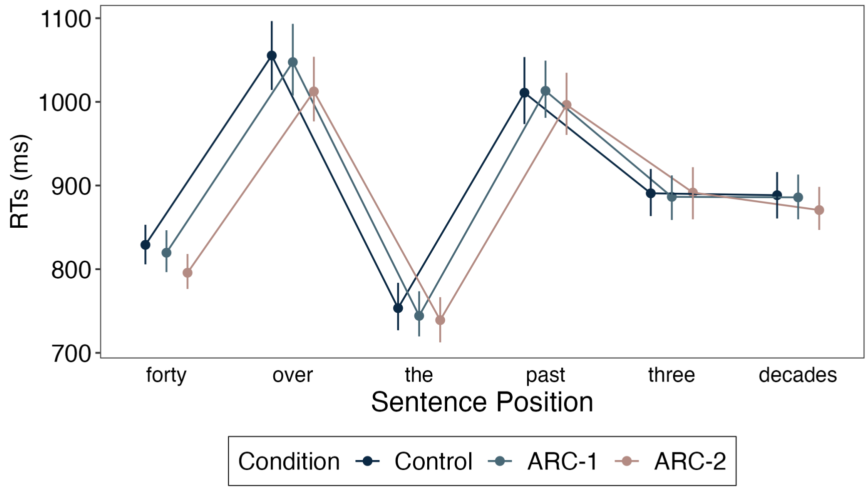


Figure 3.3: Word-by-word Maze latencies by condition for Experiment 2.

(lexical, syntactic, and semantic) as well as decision time. The task requires remembering previous sentence content, and integrating incoming material with this representation on a word-by-word basis. The Maze is advantageous for investigating the timecourse of dependency resolution, as it encourages highly incremental processing and generally produces localized results (i.e. typically at the region of interest, rather than downstream in the spillover). A number of well-established sentence processing effects have been replicated using the Maze (Forster et al., 2009; Witzel, Witzel, & Forster, 2012). Furthermore, Van Handel (2022) established the feasibility of using the Maze to investigate effects of implicit prosody.

In the current experiment, a Yes/No comprehension question was presented after half of the trials. Each experimental session took approximately 60-75 minutes.

3.3.2 Results

Results are plotted in Figure 3.3. At both the pre-critical (*forty*) and critical (*over*) regions, mean RTs reveal no difference between the CONTROL vs. ARC-1 conditions, but suggest numerically faster RTs to ARC-2. This difference may have begun to emerge prior to the critical region due to participants developing a strong expectation for lexical and structural parallelism across the clauses containing the antecedent and ellipsis site throughout the course of the experiment. The spillover regions (*the past*) suggest comparable RTs to all three conditions.

Maze latencies were analyzed using Bayesian linear mixed-effects models using the `brms` package in R (Bürkner, 2017). Because the number of words in the antecedent region, and consequently the amount of content retrieved at the ellipsis site, varied across items, the analysis utilized residual log RTs on the critical and summed spillover regions. Residual log RTs (plotted in Figure 3.4) were computed using the `lme4` package in R (Bates et al., 2015) by fitting a linear mixed-effects model to log RTs with characters per word and word position as fixed slopes and with by-participant random intercepts. `brms` models[22] were fit to latencies at the critical and summed spillover regions, using reverse-helmert contrasts, and included the maximal random effects structure, following Barr, Levy, Scheepers, and Tily (2013). The $\hat{R}$-diagnostic as well as visual inspection of posterior predictive check plots indicated model convergence (Gelman, Carlin, Stern, & Rubin, 2014). Regression weights and 95% credible intervals are provided in Table 3.17.

[22]`brm(logRT ~ Structure + (1 + Structure | Subject) + (1 + Structure | Item))`

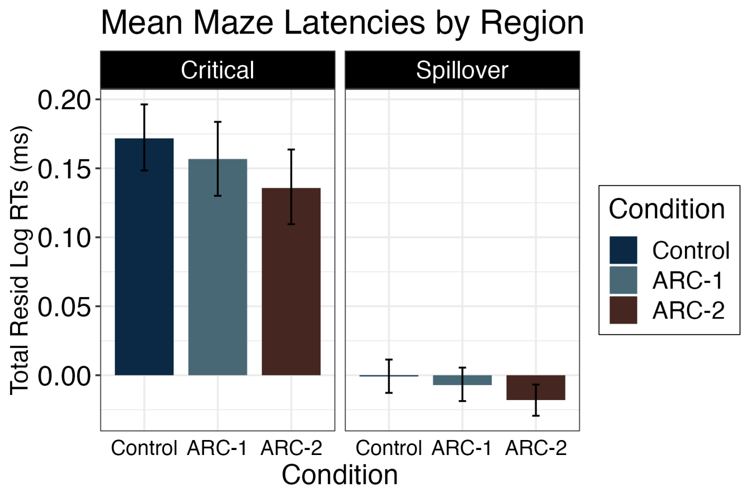


Figure 3.4: Mean residual log RTs by condition on the critical and summed spillover regions for Experiment 2.

Critical (*over*)	$\hat{\beta}$	95% CrI
C VS. ARC-1	-0.01	(-0.05,0.02)
C, ARC-1 VS. ARC-2	-0.03	(-0.08,0.02)
Spillover (*the past*)	$\hat{\beta}$	95% CrI
C VS. ARC-1	-0.01	(-0.02, 0.01)
C, ARC-1 VS. ARC-2	**-0.02**	**(-0.05,-0.0008)**

Table 3.17: Bayesian linear mixed-effects models fit to residual log RTs at the critical and summed spillover regions of Experiment 2.

Residual log RTs suggested that the data numerically trend in the right direction to support CSR (i.e., ARC-1 and ARC-2 are both faster than CONTROL) at the critical and spillover regions. However, neither the difference between CONTROL and ARC-1 ($\hat{\beta}$ = -0.01, 95% CrI = [-0.05, 0.02]) nor the difference between CONTROL and ARC-2 ($\hat{\beta}$ = -0.03, 95% CrI = [-0.08, 0.02]) was credibly different at the critical region. At the spillover region, ARC-1 was numerically faster than CONTROL, but not credibly so ($\hat{\beta}$ = -0.02, 95% CrI = [-0.02, 0.01]). The CONTROL VS. ARC-2 comparison, however, revealed a credible difference such that ARC-2 was faster ($\hat{\beta}$ = -0.02, 95% CrI = [-0.05, -0.0008]).

3.3.3 Discussion

Overall, the pattern of results was more consistent with CSR, but somewhat mixed: ARC-1 was no slower than the CONTROL condition, and ARC-2 was credibly faster, suggesting that prosodic segmentation may facilitate cue-based retrieval in the presence of a unique cue, but not for ARC-internal content. At the same time, the results do not support the predictions of Compression or Backgrounding, which would have predicted slower RTs to the ARC-1 condition compared to the CONTROL. However, the lack of difference between ARC-1 and CONTROL is surprising under CSR. Under a view that purely relies on contextual partitions in memory, we might have expected the fastest reaction times to the ARC-1 condition, because it contains the least content and should therefore engender less contextual interference, by hypothesis. This suggests that something else is at play for ARC-internal antecedents, which renders the benefit of segmentation inert.

Taken together with the conclusions from Experiment 1, we take the results to provide support against a mechanism where ARC content is degraded, like Compression. The role of Backgrounding cannot entirely be ruled out, however. The lack of credible difference between CONTROL and ARC-1 is consistent with a state of affairs where both a dispreference for ARC-internal antecedents (per Backgrounding) and a reduction in interference for prosodically segmented content (per CSR) play concurrent roles in the (post-RC) representation of sentences containing sentence-medial ARCs. In other words, prosodic boundaries may generally serve to segment content into distinct encoding contexts, rendering the content contained within the same context as a retrieved item relatively more active than the content of other contexts. At the same time, the semantic/pragmatic status of certain segments may also influence a dispreference for those contents through the encoding of a context-level +BKGRD feature, like in (41) (c.f. S. Kim & Xiang, 2022, 2024). The current experiment cannot comment on (i) whether such a preference is a consequence of the features on the retrieval probe (i.e., a search for +MAIN content), or (ii) whether it may arise from a more general dispreference for non-parallel information structure in ellipsis resolution (i.e., a dispreference for establishing a dependency with backgrounded content when currently situated within the main clause, as in (41)). Based on the discussion in §3.1.2.2, we think the second option more likely. In relation to point (i), it should be noted that the Backgrounding mechanism put forward in §3.1.3 does not assume that "main" discourse status is encoded as a feature[23]. It's possible, however, that a context-level feature like +MAIN could be incrementally encoded along with each item, and updated only in the appropriate structural or discourse context. Because the Maze encourages a higher degree of incremental commitment to interpretation than other reading tasks, it may not be an ideal task for teasing these retrieval vs. post-retrieval possibilities apart

(41) CSE + Backgrounding

1 +BKGRD 2 3

The author, | who wrote **two novels**, | met the hack that wrote **forty** _

1 : *author,* → context shift

2 : *novels,* → mark current context with +BKGRD, context shift

3 : *forty* _ → **!** difficulty retrieving *novels* from +BKGRD encoding context

If discourse-structural parallelism has an influence on ellipsis resolution, then we might expect to see a benefit for an ARC-internal antecedent emerge in a structure like (42), where a pressure for resolving ellipsis across contextually-like discourse

[23]It's not quite clear what it would mean under an SDRT framework to encode the main clause with a +MAIN feature, as this is not a proposed rhetorical relation (Asher & Lascarides, 2003), and because coherence relations are meant to capture a link between discourse segments, not the isolated status of a particular segment. In that sense, the theory does not import straightforwardly into a model of memory dependent on feature-based item encodings.

segments coupled with a benefit for prosodic segmentation could facilitate access to content within a backgrounded context.

(42) +BKGRD +BKGRD

The author, | who wrote **two novels**, | met the hack, | who wrote **forty** _

Note that Backgrounding, as it was defined in §3.1.3, is a process that applies given an appropriate discourse context once the processing of an ARC is complete. In order for context-level encoding as in (42) to obtain, then, we must further assume that a preference for discourse structural parallelism can serve to generate an expectation for a backgrounded segment in advance of encountering the right ARC boundary, and in addition, that this expectation can be used to guide the search for an NPE antecedent. In the case of (42), this would additionally suggest that a preference for parallelism may overrule the tendency to treat sentence-final ARCs as at-issue. Experiment 3 will consider this further, but the results rule out this account.

Ultimately, I suggest that the Reinstantiation mechanism introduced in §3.1.3 can account for these results and the ARC bypassing effect without positing that the retrieval mechanism itself is sensitive to something like background discourse status. I postpone a full discussion of this until Experiment 6 (§4.3).

Before turning to Experiment 3, I address some potential issues with the current experimental design. Though the RTs in the spillover region trend in the right direction to support CSR, the design did not include baseline conditions without NPE against which to compare the critical sentences. This is potentially relevant as we would ideally want to establish whether there are other differences in acceptability or reading times between these structures. For example, establishing a contrastive relationship (e.g., *TWO novels* vs. *FORty novels*) across segments with distinct discourse status may be difficult independent of resolving NPE. In addition, the items used in Experiment 2 varied in the length of the antecedent region, and the critical region itself varied in length significantly, because some of the critical prepositions and subordinators used were quite short (2 characters), and others were longer (up to 7 characters). While the current results are suggestive of CSR, it is curious that we observe a credible difference in the spillover but not the critical region, especially given that this is atypical of Maze results. As such, these length-related factors may have introduced unnecessary variability in our region of interest. Experiment 3 extends the design of Experiment 2 in order to address these concerns.

3.4 Experiment 3

The goal of Experiment 3 was to replicate and extend Experiment 2, once again contrasting the predictions of Context-Sensitive Retrieval and Backgrounding. As such, the design used identical constructions crossing Structure (RRC-RRC, ARC-RRC, RRC-ARC, ARC-ARC) with Ellipsis (+NPE, -NPE). The levels of the Structure factor are labeled to indicate the type of subject-modifying RC followed by the type of object-modifying RC. +/-NPE refers to whether there was an overt noun immediately preceding the critical region.

First, we predicted a main effect of ellipsis resolution, such that +NPE conditions should result in longer latencies than -NPE conditions due to retrieval of an antecedent from memory. Under CSR, we predicted a smaller ellipsis resolution cost (i.e., a lesser difference between RTs to +/-NPE conditions) in segmented structures (those containing an ARC) compared to the non-segmented RRC-RRC control condition. Under Backgrounding, we predicted a larger ellipsis resolution cost for ARC-internal antecedents (the ARC-RRC and ARC-ARC conditions) than RRC-internal antecedents (the RRC-RRC and RRC-ARC conditions). Additionally, Backgrounding predicts no difference in NPE resolution between the RRC-RRC and RRC-ARC condition. Under a mixture of CSR and

Hypothesis		Predictions
CSR	1	Segmented structures are faster: RRC-RRC > ARC-RRC, RRC-ARC, ARC-ARC
	2	Smaller segments are faster: ARC-RRC, ARC-ARC < RRC-ARC
Backgrounding	1	Retrieval from +BKGRD segment is slower: ARC-RRC, ARC-ARC > RRC-RRC, RRC-ARC
	2	No benefit for segmentation: RRC-RRC = RRC-ARC
CSR + Back.	1	Segmented structures are faster: RRC-RRC > RRC-ARC
	2	But retrieval from +BKGRD segment is slower: RRC-RRC = ARC-RRC
	3	And discourse parallelism helps: ARC-ARC < RRC-RRC

Table 3.18: Predictions for Experiment 3 under CSR, Backgrounding, and CSR + Backgrounding. (In)equalities indicate how NPE cost is predicted to vary by Structure.

Backgrounding, we predicted a replication of the results of Experiment 2: that NPE cost should not differ meaningfully between the RRC-RRC and ARC-RRC conditions, but should be credibly smaller for RRC-ARC compared to RRC-RRC/ARC-RRC. Crucially, the mixture hypothesis additionally predicts the fastest NPE resolution for the ARC-ARC condition, under the additional assumption that syntactic and discourse parallelism can serve to generate an expectation for a +BKGRD segment in advance of its completion. These predictions are summarized in Table 3.18.

3.4.1 Method

3.4.1.1 Participants

120 Prolific participants took part in the experiment and were paid $12/hr. Participation restrictions were the same as in Experiment 2. Any participant who failed more than 50% of Maze trials or answered less than 70% of the comprehension ques-

Structure	Sentence **critical region** | spillover region
RRC-RRC	The delusional billionaire that owned five houses pitied the honest butcher that owned zero (houses) **despite** many years of saving.
ARC-RRC	The delusional billionaire, who owned five houses, pitied the honest butcher that owned zero (houses) **despite** many years of saving.
RRC-ARC	The delusional billionaire that owned five houses pitied the honest butcher, who owned zero (houses) **despite** many years of saving.
ARC-ARC	The delusional billionaire, who owned five houses, pitied the honest butcher, who owned zero (houses) **despite** many years of saving.

Table 3.19: Example item set from Experiment 3. -NPE conditions included the nominal in parentheses.

tions accurately was excluded. 13 participants were excluded on this basis, leaving the remaining 107 participants. In order to ensure balanced observations across experimental lists, 13 additional undergraduate students from the UC Santa Cruz subject pool were recruited to replace the 13 excluded from Prolific. Participants recruited through the subject pool were compensated with course credit. The final analysis included data from 120 participants.

3.4.1.2 Materials

64 item sets (Table 3.19) were constructed, using the same structural configurations as in Experiment 2, with the addition of the ARC-ARC structure. In order to address some of the design issues raised in the discussion of those results, items were modified such that the object of the first RC always contained exactly two words (a numeral and noun), and the critical region was always between 4-8 characters.

Experimental items were counterbalanced across 8 Latin-squared lists and were randomly presented along with 42 fillers. Fillers were similar in nature to those used in Experiment 2.

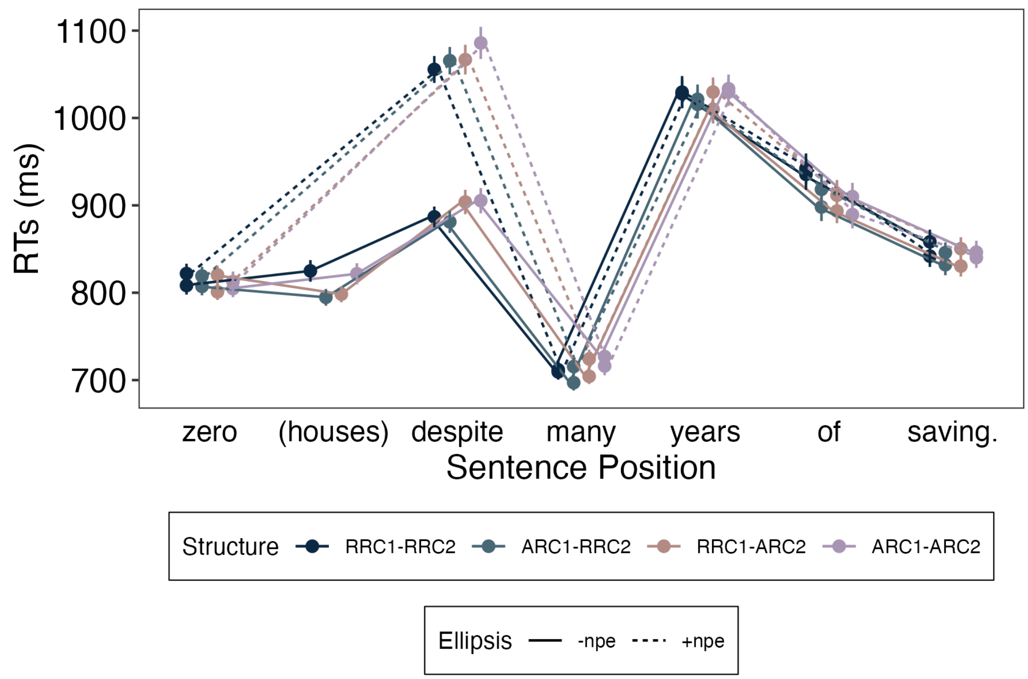


Figure 3.5: Word-by-word Maze latencies by condition for Experiment 3.

3.4.1.3 Procedure

The procedure was identical to Experiment 2, except that Experiment 3 included a comprehension question after every sentence. Each experimental session took approximately 50-60 minutes.

3.4.2 Results

brms linear mixed-effects models[24] were fit to log RTs at the critical and spillover regions and included the maximal random effects structure (Barr et al., 2013). The $\hat{R}$-diagnostic and posterior predictive checks indicated good model convergence (Gelman et al., 2014). Results are plotted in Figure 3.5 and model results are in in Table 3.20.

I discuss the effects at the critical region first. We observed a main effect of El-

[24]`brm(logRT ~ Structure*Ellipsis + (1 + Structure*Ellipsis | Subject) + (1 + Structure*Ellipsis | Item))`

Effect	Critical (*despite*) $\hat{\beta}$	95% CrI	Spill1 (*many*) $\hat{\beta}$	95% CrI	Spill2 (*years*) $\hat{\beta}$	95% CrI
NPE	**0.17**	**(0.14,0.2)**	0.01	(-0.02,0.03)	0.01	(-0.02,0.04)
ARC-RRC	-0.001	(-0.2,0.2)	-0.001	(-0.02,0.01)	-0.002	(-0.02,0.02)
RRC-ARC	0.01	(-0.01,0.03)	0.01	(-0.01,0.03)	-0.001	(-0.02,0.02)
ARC-ARC	0.01	(-0.01,0.03)	0.01	(-0.01,0.03)	0.01	(-0.02,0.03)
NPE*ARC-RRC	0.02	(-0.02,0.07)	0.01	(-0.02,0.05)	-0.002	(-0.05,0.04)
NPE*RRC-ARC	-0.01	(-0.06,0.03)	0.01	(-0.03,0.04)	0.001	(-0.04,0.05)
NPE*ARC-ARC	0.01	(-0.04,0.05)	-0.02	(-0.06,0.02)	-0.001	(-0.05,0.04)

Table 3.20: Bayesian linear mixed-effects models fit to log RTs at the critical and spillover regions of Experiment 3.

lipsis at the critical region such that RTs to +NPE conditions were slower than RTs to -NPE conditions ($\hat{\beta}$ = 0.17, CrI = [0.14, 0.2]). No credible effects of structure emerged. Among the three conditions present in Experiment 2, the pattern of results observed in the current experiment was numerically consistent with that of Experiment 2. The difference in +/-NPE conditions was smaller for in the RRC-ARC condition compared to the RRC-RRC and ARC-RRC conditions. Model results revealed that this trend was not credible ($\hat{\beta}$ = -0.01, CrI = [-0.06, 0.03]). Similarly, there was no meaningful difference of differences between the RRC-RRC and ARC-ARC conditions ($\hat{\beta}$ = 0.01, CrI = [-0.04, 0.05]). No credible effects emerged at either spillover region.

Because visual inspection of the results suggested that differences in RTs among the structures emerged at the pre-critical region and the penultimate region of the sentence, post-hoc analyses were conducted on these regions. `brms` models[25] were fit to log RTs at each region. These results are reported in Table 3.21.

At the pre-critical region (*houses*), RTs to the ARC-RRC and RRC-ARC conditions were faster than RTs to the RRC-RRC and ARC-ARC conditions. We hypothesized that the pre-critical region was a potential locus of resolving information structure in the

[25] `brm(logRT ~ Structure + (1 + Structure | Subject) + (1 + Structure | Item))`

Effect	Pre-critical (*houses*) $\hat{\beta}$	95% CrI	Spill3 (*of*) $\hat{\beta}$	95% CrI
ARC-RRC	-0.02	(-0.05,0.004)	-0.02	(-0.04,0.002)
RRC-ARC	-0.01	(-0.04,0.02)	**-0.02**	**(-0.04,-0.001)**
ARC-ARC	0.001	(-0.02,0.03)	**-0.03**	**(-0.05,-0.01)**

Table 3.21: Post-hoc analysis for Experiment 3. Bayesian linear mixed-effects models fit to log RTs at the pre-critical and penultimate regions.

-NPE conditions, in that it may have cued contrast between *FIVE houses* in the first RC and *ZEro houses* in the second RC. Under this explanation, we expected to see greater difficulty for structures where the RCs were distinct in terms of discourse status (i.e., MAIN/BKGRD or BKGRD/MAIN). This is the opposite of the pattern actually observed; instead, the results exhibit faster RTs to these conditions. Therefore, it may be that structures with greater contextual dissimilarity incur a processing benefit, due to reduced load during encoding. However, the model results showed no credible differences between structures, so we discount this possibility.

At the penultimate region of the sentence (*of*), the RT pattern reflects a penalty for the non-segmented structure (RRC-RRC) compared to the rest. We hypothesized that this, too, could be due to a benefit for segmented structures, as a result of reduced memory load. Model results indicated no credible difference between the RRC-RRC vs. ARC-RRC conditions ($\hat{\beta}$ = -0.02, CrI = [-0.04, 0.002]), but did result in credible differences between RRC-RRC vs. RRC-ARC ($\hat{\beta}$ = -0.02, CrI = [-0.04, -0.001]) and RRC-RRC vs. ARC-ARC ($\hat{\beta}$ = -0.03, CrI = [-0.05, -0.01]). It's somewhat mysterious that the RRC-RRC vs. ARC-RRC difference was non-credible, as this is one structural configuration in which ARC discounting effects are typically observed in offline measures. Nevertheless, we suggest that the effect of Structure here may reflect (i) a late segmentation benefit, per CSE, independent of dependency resolution, or (ii) reduced contextual interference within the current segment (i.e., in the presence of a final ARC).

3.4.3 Discussion

Overall, Experiment 3 failed to find convincing support for either CSR or Backgrounding, as the results revealed no credible support for the predicted interactions under either account. The predictions of CSR, that NPE cost for segmented structures is reduced and that NPE cost for segments with less content is reduced, were not met. The critical prediction of Backgrounding, that NPE cost for ARC-internal antecedents should be greater, was similarly not met. The data trend in a direction partially consistent with CSR in that NPE cost for RRC-ARC was numerically smaller than that of RRC-RRC, consistent with prediction 1 of CSR + Backgrounding in Table 3.18, albeit not meaningfully so. In addition, there was no difference between the RRC-RRC and ARC-RRC conditions, which aligns with prediction 2 of CSR + Backgrounding. Prediction 3, that NPE cost for ARC-ARC is less than for RRC-RRC, was not met. However, this prediction was contingent on the assumption that the parser may track an expectation for a +BKGRD segment prior to its completion. Given the reliance on this additional assumption, it would be reasonable to assume that prediction 3 does not provide crucial evidence for a combined CSR + Backgrounding account. Therefore, the possibility that both mechanisms are concurrently active remains.

Interestingly, the facilitation of RTs for segmented structures in the penultimate region did implicate a role for Context-Sensitive Encoding, although the results did not validate the role of Context-Sensitive Retrieval. I discuss this further in the General Discussion.

3.5 General Discussion

The conclusions of Experiments 1-3 are summarized in Table 3.22.

The current chapter has established that (i) linguistic memory is sensitive to con-

Experiment	Hypotheses	Findings
Exp. 1	✔ CSE ✗ Compression	Syntactic memory for ARC content is retained better than for RRC content.
Exp. 2	✔ CSR ? Backgrounding	Benefit for retrieval from segmented sentences. ARC-internal access is not facilitated/inhibited.
Exp. 3	✔ CSE ? CSR ? Backgrounding	Late benefit for segmented sentences. Non-meaningful benefit for retrieval from segmented sentences. ARC-internal access is not facilitated/inhibited.

Table 3.22: Summary of findings from Experiments 1-3.

textual distinctions (per CSE) but (ii) the mechanism underlying the formation of contexts in memory does not render certain contexts inaccessible to the parser's operations online (contra Compression, and possibly, contra Backgrounding). This was evidenced by two facts about the memory representation of appositive relative clauses: (i) that while appositive content is both discounted during parsing and better retained in memory, and so must be contextually demarcated as a distinct unit in the memory representation of a sentence, (ii) it is not the case that appositive-internal content is less accessible than main clause content.

The critical prediction of the Context-Sensitive Retrieval (CSR) hypothesis was that the contextual interference that results from successful retrieval of an item should be reduced for segmented structures. The data suggestively hint at this pattern. Specifically, the trend that RRC-ARC NPE cost was smaller than RRC-RRC NPE cost implicated the role of CSR, but only weakly, as this difference was meaningful in Experiment 2 but not in Experiment 3. Taken together, the results of Experiments 2 and 3 suggest that any effect of contextual interference arising from cue-based retrieval is very small if present at all. Experiment 3 did, however, provide additional support for Context-Sensitive Encoding, due to the later facilitation of RTs for segmented structures (in the penultimate region of the sentence, 3 words past the critical region). I conclude that contextual partitions drive encoding, but it's likely that item-level features outweigh temporal context features during retrieval.

Alternatively, it may be that the effect of contextual interference is not observable during dependency resolution because it is hypothesized to arise as a consequence of successfully retrieving an item, but does not guide the search process itself. Moreover, the assumption of CSR that retrieval of an item reinstates that item's context may be difficult to observe in sentence contexts because the elements of sentences are inherently interconnected. Cue-based retrieval processes in memory have been argued to underlie the formation of a variety of linguistic dependencies, and so, comprehenders may often be reinstating previous contexts in order to carry out normal parsing operations. This is, of course, one of the properties that sets sentence memory apart from list memory: that interpreting sentences requires regularly forming dependencies between non-adjacent elements. As such, it may generally be difficult to observe an effect of CSR during sentence comprehension.

More generally, though, the current chapter has established the effect of Context-Sensitive Encoding. We've shown that ARC boundaries lead to better retention of syntactic content in memory. This state of affairs provides evidence against the particular hypothesis proposed by Dillon et al. (2017) that the surface-level details or syntactic features associated with appositive content are compressed in memory and are therefore be rendered less accessible than main clause content. In addition, the ARC bypassing effect evidences that the presence of appositive relative clause boundaries facilitate a comprehender's ability to navigate between linguistic contexts. This suggests that during the processing of a sentence with an ARC, comprehenders are able to make use of contextual cues to segmenthood in order to bypass intervening parenthetical content. I return to a discussion of this, in the context of the Reinstantiation mechanism, in Chapter 4 (§4).

Despite ample evidence for discounting and bypassing, previous work has been generally unsuccessful in identifying the particular linguistic source of the benefit for appositives. As discussed at length in §3.1.2, a number of previously entertained discourse-based explanations have failed to account for the processing profile of appositives: it has been established that their independence in terms of speech acts (Duff et al., 2023), discourse units (Duff et al., 2023), and relation to the primary QUD of a discourse (Kroll & Wagers, 2019) are inadequate in explaining their behavior. Some evidence supports the role of prosodic segmentation in discounting; recall that Kroll and Wagers (2019) found that the difference in acceptability ratings between ARCs and RRCs collapses when RRCs are visually chunked using line breaks. However, it is possible that visually chunking a sentence with an RRC such that the relative clause content occupies a distinct line may cause readers to be more likely to treat that content as an aside, regardless of the linguistic structure. On the other hand, direct discourse, which should presumably also be phrased independently in a reader's implicit prosodic representation (Yao et al., 2011; Yao & Scheepers, 2011), did not show discounting relative to indirect discourse (Duff et al., 2023). Thus, the role of prosodic segmentation in driving the discounting effect remains speculative for the time being, but Experiment 6 solidifies this intuition.

In sum, previous work has only managed to find evidence of discounting and bypassing for appositive relative clauses and nominal appositives. The idea that these effects are due to a special property of appositives in particular would be unexpected under an account like CSE, which assumes their ability to be discounted arises from more general linguistic features (i.e., the presence of sentence-medial prosodic boundaries). The experiments in Chapter 4 (§4) set out to address this puzzle.

Chapter 4

Coordinates in Memory

In certain respects, Chapter 3 left off with more questions than answers: it remains unclear why appositives, but not other similar independent units, display discounting and bypassing effects. This chapter aims to determine whether such processing independence is specific to appositives, or if instead, its source can be pinned on prosodic independence more generally. To that end, we investigate other prosodically independent constructions that do not hold a subordinate discourse status. One possibility we considered in Chapter 3 was that the backgrounded discourse status of ARCs obscured the ability to observe an effect of Context-Sensitive Retrieval (CSR), because CSR and Backgrounding canceled each other out. If the ability to discount/bypass is a general property of more-segmented linguistic structures, we should expect to find evidence that other structures containing medial boundaries facilitate contextual partitioning in memory, and in turn aid cue-based retrieval during the resolution of item-to-item dependencies. In order to probe this question, the current chapter investigates the processing of two types of focus-sensitive coordination constructions, which contain prosodically independent units.

4.1 Experiment 4

Experiment 4 investigates noun phrase ellipsis resolution across the sentence-medial boundary of focus-sensitive coordination constructions that prosodically separate their coordinates, as in (1).

(1) Imala met not only the inventor with **fifty gizmos**, but also the one with **three** <**~~gizmos~~**> after the collaboration last winter.

Note that this structure is an at-issue analogue of the sentence-final ARC condition where we observed tentative evidence for Context-Sensitive Retrieval in Experiment 2 of the previous chapter. Proposed encoding contexts under CSE/CSR for ARCs and *not only...but also* structures (NOBA) are represented schematically in Table 4.1. In part, the current experiment makes another attempt to assess the role of Context-Sensitive Retrieval across prosodic phrases that are in a coordinating discourse relationship (linked via a Contrast relation). If sentence-medial segmentation facilitates cue-based retrieval operations (given a clear cue to the antecedent), noun phrase ellipsis resolution should be less costly in NOBA structures compared to *and also* (AA) structures, a focus structure-identical counterpart to NOBA without prosodic separation between coordinates. The proposed contextual representation for AA is also provided in Table 4.1.

The current experiment contrasts the predictions of Context-Sensitive Retrieval with Visibility (Carlson et al., 2009; Harrington Stack & Watson, 2023; Schafer, 1997; Van Handel, 2022), introduced in §2.2.1 and §2.3.3. Previous work on prosodic parsing suggests that the presence of an earlier prosodic boundary can inhibit access to pre-boundary content. Recall that one interpretation of the Visibility hypothesis suggests that dependency resolution across an intervening prosodic boundary should be more difficult than in the absence of a previous boundary (see Carlson et al., 2009). §2.3.3

	Contextual Representations under CSR
	Experiments 2-3
ARC	The author that wrote **two novels** met the hack, who wrote **forty** _
RRC	The author that wrote **two novels** met the hack that wrote **forty** _
	Experiment 4
NOBA	not only the inventor with **fifty gizmos**, but also the one with **three** _
AA	the inventor with **fifty gizmos** and also the one with **three** _

Table 4.1: Proposed contextual representations under Context-Sensitive Retrieval for the structures used in Experiments 2/3 vs. Experiment 4.

argued that this can be captured under a temporal context mechanism by assuming that all else equal, the contents of the current or most recent encoding context are most active, and accessing the contents of a previous segment requires reinstatement of that segment's context. I termed this approach CSE-Visibility, repeated in (2).

(2) **CSE-Visibility**: The contents of the current encoding context (i.e., the current prosodic phrase) are most accessible, because they partially match with the current contextual state. Retrieving the contents of an earlier encoding context requires reinstantiation of an earlier contextual state.

Under CSE-Visibility, ellipsis resolution should be more costly in NOBA structures than in AA structures. This is represented in (3). This prediction contrasts directly with that of Context-Sensitive Retrieval.

(3) Contextual representation for *not only...but also* under CSE-Visibility

not only the inventor with **fifty gizmos**, but also the one with **three** _

We consider the role of CSE-Visibility for this experiment because Visibility effects have been argued to stem from a fundamental property of prosodic phrase boundaries. If the predictions of Visibility are borne out in the current experiment, the fact that Visibility does not apply to ARC boundaries must be explained. This point is elaborated upon in the context of the results. Before discussing the predictions in more detail, the following section first introduces relevant background information on linguistic focus, then turns to a review of work on the structure and processing of focus-sensitive coordination.

4.1.1 *not only...but also*

Experiment 4 tests the resolution of noun phrase ellipsis across the conjuncts of *not only...but also* (NOBA) constructions (4), which involve prosodic separation of the first and second coordinates, compared to *...and also...* (AA) constructions (5), which can but need not prosodically separate their coordinates. Both constructions involve focus-sensitive coordination, along with coordinators like *let alone*, *much less*, *as well as*, *either...or*, and others (De Vries, 2005; Hulsey, 2008; Toosarvandani, 2010; Wu, 2022). These constructions are sometimes argued to be structurally distinct from ordinary coordination (Hulsey, 2008), and involve clausal coordination, obligatory ellipsis (but see Toosarvandani for counterarguments), and the presence of contrastive foci bearing pitch accents within each coordinate.

(4) Tara sails not only [DINghies]$_{\mathrm{F}}$, but also [KEELboats]$_{\mathrm{F}}$.

(5) Tara sails [DINghies]$_{\mathrm{F}}$ and also [KEELboats]$_{\mathrm{F}}$.

Broadly, *linguistic focus* is a grammatical device used to signal contrast through evoking alternatives. Under one formal theory of focus[26], Alternative Semantics (Rooth, 1985, 1992b), focus interpretation is constrained via a "squiggle" operator ($\sim$), the contribution of which is presuppositional. The theory rests on the assumption that a focused constituent presupposes a set of contextually salient alternatives, resulting in both an ordinary semantic value (6a) as well as a focus semantic value (6b). The squiggle establishes a relation between a covert contextual variable p_1 and the phrase it adjoins to, such that the value of p_1 is a subset of the focus semantic value of the phrase and also contains the ordinary semantic value of the phrase (6b).

(6) [TARA$_F$ sails] $\sim p_1$ ($\rightsquigarrow$ not Margot, Ramona, Finn…)

a. $[\![$Tara sails$]\!]^o$ = [$\lambda x.x$ sails](Tara) = 1 iff Tara sails

b. Presupposition:

$p_1 \subseteq [\![$TARA$_F$ sails$]\!]^f$ = {$[\![$sails$]\!]^f(x)$ | $x \in$ D$_e$} = {Tara sails, Margot sails, ...}

Contrast may be signaled through a variety of linguistic means: via the discourse context of a sentence (7), its structure (8), certain focus-sensitive particles like *only* or *even* (9), or prosodic emphasis via the presence of a focal pitch accent (10). Throughout the current section, I will indicate focused content via [ ]$_F$ and the position of an accented syllable within a focus using SMALL CAPS.

(7) A: *What* does Tara sail?

B: Tara sails [a *DINghy*]$_F$

(8) It was [a DINghy]$_F$ that Tara sailed.

(9) Tara sailed only [a DINghy]$_F$

(10) Tara sails [a DINghy]$_F$

a. $\rightsquigarrow$ not {a schooner, a ketch, a catamaran,...}

[26]There are others (Klein & Von Stechow, 1982; Krifka, 1992), but a full discussion would take me too far afield, so I set them aside here.

In examples (7)-(9), multiple linguistic cues jointly signal the presence of focus on the DP *a dinghy*. In (7), for example, the *wh*-element in the preceding discourse context corresponds with the position of the focus, which also bears a focal pitch accent. Regardless of the particular method of signaling focus (by focus-sensitive operator, discourse context, or syntactic structure), there is a common interpretive effect: to evoke alternatives to the focused content. The presence of focus allows comprehenders to arrive at the understanding that it was a dinghy that was sailed, in contrast to a schooner, a ketch, a catamaran, or some other sailing vessel (10a). Note that the size of the focus and the position of the pitch accent are dissociable. Consider the mini-discourse in (11), where the particle *only* marks the scope of the focus. In this example, alternatives are generated over all the focus-marked content, not only the word bearing the pitch accent. Possible alternatives for this example are given in (11a).

(11) A: Tara read only [a book about DINghies]$_F$

B: No, Tara read only [an ARTicle about dinghies]$_F$

a. ⇝ not {a book about keelboats, a book about schooners, an article about keelboats, an article about schooners...}

Furthermore, a particle in a broad focus position leads to ambiguous focus structure, as in (12).

(12) Tara only read a book about dinghies.

a. ...only [READ]$_F$ a book about dinghies. ⇝ didn't memorize it

b. ...only read [a BOOK]$_F$ about dinghies. ⇝ not an article

c. ...only read a book about [DINghies]$_F$. ⇝ not keelboats

Importantly for present purposes, NOBA (13) and AA (14) are both focus-sensitive constructions that offer a minimal pair with identical syntactic and information structure, but differ in terms of (i) the presence of a focus particle marking the left edge of the first coordinate in NOBA, and (ii) the presence and strength of an obligatory prosodic boundary in NOBA (see Wu, 2022, for a similar argument about the prosodic structure of *either...or* constructions, another type of focus-sensitive coordination). I return to a discussion of the prosodic differences between NOBA vs. AA shortly, after introducing relevant syntactic and semantic properties.

(13) Ramona met *not only* [the ACtor]$_F$, *but also* [the diRECtor]$_F$.

(14) Ramona met [the ACtor]$_F$ *and also* [the diRECtor]$_F$.

Evidence for their status as focus-sensitive coordination (FSC) comes from the fact that both constructions diverge from syntactic characteristics of ordinary coordination. As such, they are argued to involve focus-sensitive ellipsis (FSE) across propositional coordinates containing contrastive foci (Hulsey, 2008), as has been argued for other focus-sensitive coordination constructions (Wu, 2022). Focus sensitive coordination does pattern like ordinary coordination in certain respects (see Hulsey, 2008, for relevant diagnostics), but in contrast to standard coordination, FSC may not coordinate bare NPs (15). Hulsey therefore argues that while FSC may sometimes appear to coordinate smaller constituents, it must always involve coordination of propositions (c.f. Toosarvandani, 2013), and that smaller surface syntax is subject to usual constraints on ellipsis.

(15) a. John might paint a chair and (an) armoire.
     b. *John might paint not only a chair, but also armoire.
     c. *John might paint a chair and also armoire.

Hulsey's (2008) proposed structure for FSC involves raising of both foci followed by ellipsis, leaving a remnant (*director*) and correlate (*actor*) in each coordinate, respectively, which must be associated in order to establish contrast between foci; this is exemplified in (16).

(16) Ramona$_1$ met not only t_1 [the ACtor]$_{F,2}$ <~~t_1 meet t_2~~> , but also t_1 [the diRECtor]$_{F,3}$ <~~t_1 meet t_3~~>.
correlate = *actor*; remnant = *director*

In order to use these constructions to probe the role of prosodic boundaries in retrieval, it is necessary to understand something about their processing profile. First, I note that comprehenders generally attend to focus incrementally. The assignment of focus structure and accent position have been shown to proceed in real-time during sentence comprehension (e.g., Baumann & Schumacher, 2020; S. Birch & Rayner, 1997). Readers also display online sensitivity to the presence of alternatives in context (Hoeks, 2023). Together, these facts suggest that the processing of focus is a multi-stage process. §4.2.5 will consider how more general principles of focus processing may interact with the results of Experiments 4-5; I postpone a discussion of this until then. Processing of FSC in particular constitutes a unique case, because (i) its focus-sensitivity is sometimes not determined immediately (e.g., in the absence of an initial coordinator), and (ii) it involves resolution of focus-sensitive ellipsis. These properties are discussed at length throughout the remainder of this section, but I begin with some basic facts about the processing of FSC.

The majority of previous work on the processing of FSC concerns *much less* and *let alone* constructions (Carlson & Harris, 2018; Harris, 2016, 2023; Harris & Carlson, 2016, 2018). The authors use these FSC constructions in order to determine the default focus structure of the first coordinate, and expectations for the category of the remnant in the second coordinate. Typically, studies show penalties for establishing a focus structure where the correlate is not located in the final position of the first coordinate. This is the most common focus structure present in both corpus studies (Harris & Carlson, 2016, 2018) and Cloze tasks (Harris, 2016). NP remnants contrasting with a final correlate within the first coordinate are least costly; they yield higher acceptability judgments in listening, faster SPR latencies, and less pupil dilation during listening. In ambiguous constructions, the presence of a focal accent on a non-local potential correlate can reduce, but not entirely eliminate, the penalty for establishing non-final contrast. For example, Harris and Carlson (2018) investigate whether the position of an early focal accent, like in (17), can override the default preference for the local correlate, but nevertheless find a preference for *the governor*. Thus, they propose that the default parsing preference for locating the correlate in such constructions is to "keep it local and final" (Harris & Carlson, 2016). This replicates across a number of studies.

(17) John didn't write an ARticle that exposed the governor, let alone {the president/a book}.

Overall, this may suggest that default assumptions about the position of focus generally take precedence over accent position in the incremental assignment of focus structure. Put differently, the earliest cue to focus-sensitive coordination (in the case of (17), this is the presence of the *let alone* coordinator) generates an expectation for a local correlate, regardless of previous accent location. Assuming that the default structure is confirmed (where the remnant corresponds to *the president* in the case of (17) above), and accent assignment on *president* subsequently proceeds, the preferred structural position of the correlate and the availability of a contrasting element should generally take priority over the presence of an earlier accent on *ARTicle*. Under this view, the properties of the remnant along with default structural expectations about the syntactic position of focus guide the search for the correlate, rather than the accessibility or prominence of previously encountered content.

A related notable property about the processing of FSC is that the search for a correlate is strongly guided by syntactic parallelism. Carlson and Harris (2018) investigate the processing of adjectival contrasts with *much less* ellipsis, and find that ADJ

+ N correlates are preferred given an ADJ + N remnant in examples like (18).

(18) I don't own a PINK hat, much less a RED one.

(19) I don't own a (pink) hat, much less a red one.

Carlson and Harris find a dispreference for zero-adjective contrasts (19), where the first coordinate lacks the adjective *pink*, in listening, completion studies, and self-paced reading. They attribute this to the fact that the parser prefers to search for a syntactically parallel remnant. In example (19), note that the presence of *one*-anaphora in the second coordinate requires accent placement like in (18), where the anaphor is deaccented, and the preceding adjective bears the focal accent. Determining the appropriate focal accent position (in this case the adjective *red*) and resolving this anaphora during the initial parsing of the remnant may also facilitate the search for the ADJ + N correlate. Importantly, the authors argue that the pressure for syntactic parallelism is noteworthy given that *much less* contains an additional scalar component of meaning: it specifies that the second focus is a less likely option than the first. Carlson and Harris argue that in terms of computing a scalar relationship between the foci, a zero-adjective contrast like *not a hat* $\rightarrow$ *not a red hat* should be easier to compute than a scalar relationship like *not a pink hat* $\rightarrow$ *not a red hat.* Despite this, the preference for a parallel correlate persists.

As such, Carlson and Harris (2018) assume the tasks of the processor in focus-sensitive coordination structures are as in (20).

(20) Tasks of the processor for resolving focus sensitive ellipsis:

a. Parse the remnant (*a red one*)

b. Use syntactic structure of the remnant to locate the correlate (*a pink hat*)

c. Reconstruct ellipsis site: *much less [a red one]*$_{t1,\,F}$ *<own t*$_1$*>*

d. Establish a scale

Given the discussion of focus assignment and accent placement above, we amend their list of proposed tasks to include two additional steps. The ordering below is merely a suggestion, because we do not at present have empirical evidence to support the ordering of steps (c) and (d) with respect to other processes.

(21) Tasks of the processor for resolving focus sensitive ellipsis:

a. Parse the remnant (*a red one*)

b. Use the structure of the remnant to locate the correlate (*a pink hat*)

c. Establish the contrast set (ADJ *hat*)

d. Establish the position of the focal pitch accent (a PINK hat/a RED one)

e. Reconstruct ellipsis site: *much less [a red one]*$_{t1,\,F}$ <~~*own t*$_{1}$~~>

f. Establish a scale (only for scalar coordinators like *let alone*, *much less*)

We take all of these processes to be relevant to the processing of NOBA (used in Experiment 4) and *as well as* (used in Experiment 5) except for (f), as there is no evidence of a necessary scalar relationship between the foci in these constructions.

To our knowledge, only one study has directly investigated the processing of NOBA (Lowder et al., 2021), and an earlier, related study investigated the processing of *not* X*, but rather* Y constructions (Lowder & Ferreira, 2016). In a visual world study, Lowder et al. (2021) directly compare the difference between NOBA vs. AA in terms of establishing contrasting alternative sets using items like in Table 4.2.

Here, the authors assume that in the presence of a focus sensitive operator like *not only*, which marks an upcoming coordinate and allows comprehenders to start generating predictions about potential alternatives in advance, the penalty for an unpredictable N2 should be ameliorated relative to AA, which does not mark focus structure in advance of encountering N1. Their results confirm that the process of establishing contrast is faster for NOBA compared to AA, which they assume to be a “neutral”

	Predictable	Unpredictable
NOBA	The wedding is about to start, but we are still looking for not only the **bride** but also the **groom**...	The wedding is about to start, but we are still looking for not only the **bride** but also the **priest**...
AA	The wedding is about to start, but we are still looking for the **bride** and also the **groom**...	The wedding is about to start, but we are still looking for the **bride** and also the **priest**...

Table 4.2: Example item set from Lowder et al. (2021).

focus construction (c.f. Hulsey, 2008), at least during the processing of NP1, because there is no cue to the focus structure in advance. Their results suggest that there are meaningful differences between NOBA and AA at least in stage (21c) above. At present, we don't know whether the processing of these constructions differs at other stages as well.

In sum, the processing of focus-sensitive coordination/ellipsis is complex – it involves a number of processes across multiple levels of linguistic representation that are dependent on one another. The current study attempted to control for a number of these variables by comparing two constructions that hold many of these processes constant, and differ across only two dimensions: (i) pre-first coordinate focus particle presence and (ii) prosodic boundary strength/presence. Furthermore, the purpose of this study was to examine the role of an intervening prosodic boundary in the resolution of item-to-item dependencies, rather than in the resolution of correlate-remnant pairs or the reconstruction of focus-sensitive ellipsis itself. Thus, we used noun phrase ellipsis dependencies that involved accessing an antecedent contained within the correlate in order to probe the role of prosodic boundaries.

It is worth mentioning that AA may intuitively be preceded by a prosodic boundary as well, the presence of which is likely conditioned by the length of the first coordinate, as the likelihood of a boundary increases along with length of a particular prosodic phrase. This means that in some cases, the prosodic structure of NOBA and AA may eventually end up being identical (i.e., in the sentence-final implicit prosodic representation). Therefore, I propose the following potential first-pass and final prosodic representations for NOBA vs. AA:

(22) First/final-pass prosodic parse for NOBA

($_{\varphi}$Ramona met not only the actor) ($_{\varphi}$but also the director)...

(23) First-pass prosodic parse for AA

($_{\varphi}$Ramona met the actor and also the director)

(24) Final-pass prosodic parses for AA

a. ($_{\varphi}$Ramona met the actor and also the director)

b. ($_{\varphi}$Ramona met the actor) ($_{\varphi}$and also the director)

The important points for present purposes are that (i) AA does not require a prosodic boundary in the final-pass, whereas NOBA does, and (ii) per (23), it is unlikely that a pre-*and also* boundary would be posited in the first-pass parse, because this presumably requires information about the length of the second coordinate that is not immediately available and comprehenders are unlikely to actively predict the prosodic boundary location in the absence of a preceding cue to a FSC construction. In the default case, this boundary location should coincide with the position of focal accent. Thus, an initial coordinator like *not only* signals both focus and a prosodic boundary in advance, but a coordinator like *and also* does not (De Vries, 2005).

Notably, the stimuli in Experiment 4 contain relatively long coordinates, potentially increasing the likelihood of a boundary in the pre-coordinator position, like in (25b).

(25) Final-pass parses for long AA structures

a. ($_{\varphi}$Ramona met the author with two manuscripts and also the author with four manuscripts)

b. ($_\varphi$Ramona met the author with two manuscripts) ($_\varphi$and also the author with four manuscripts)

We therefore assume that the implicit prosodic representation for AA should be more variable than that of NOBA, because the focus particle in the latter case should allow readers to predict the presence of an upcoming prosodic boundary prior to encountering the *but also* coordinator. Furthermore, NOBA conditions included an explicit orthographic cue to a prosodic boundary (a comma), but AA conditions excluded this cue. Context-Sensitive Encoding, as stated in §2.3.3, requires that there should be relatively strong evidence for a prosodic boundary at the time of initial encoding in order for the temporal context mechanism to shift the vector prior to coordinate two in the initial parse. The mechanism does not equate contextual partitions with the final prosodic parse of a sentence, and does not specify how encoding context boundaries may be altered at a later stage. Therefore, it should predict that the expectation for an upcoming prosodic boundary in the presence of *not only*, along with the presence of a comma marking the right edge of the first coordinate, should contribute to a greater likelihood of context-shifting following the first coordinate of NOBA than AA. I return to a more in-depth discussion of these prosodic considerations in the context of the experimental results.

4.1.2 Method

4.1.2.1 Participants

55 undergraduate students from the University of California, Santa Cruz Linguistics Department Subject Pool participated in the experiment. All participants began learning English before or around the age of 6 and were compensated for their participation with course credit. Each experimental session took 45-60 minutes. 7 participants were excluded from the dataset due to low comprehension accuracy (less than 70% correct

Structure	Sentence **critical region** | <u>spillover region</u>
NOBA	Imala met not only the inventor with fifty gizmos, but also the one with three (gizmos) **after** <u>the collaboration</u> last winter.
AA	Imala met the inventor with fifty gizmos and also the one with three (gizmos) **after** <u>the collaboration</u> last winter.

Table 4.3: Example item set from Experiment 4. -NPE conditions included the nominal in parentheses. ***Key:*** NOBA = *not only* X*, but also* Y, AA = X *and also* Y

.

responses) on experimental items and fillers. The final analysis included data from 48 participants.

4.1.2.2 Materials

48 item sets were constructed, each consisting of 4 conditions. The experimental design crossed Structure (NOBA, AA) with Noun Phrase Ellipsis (+NPE, -NPE). The Structure factor varied whether the sentence contained a *not only* X*, but also* Y coordinate structure or an X *and also* Y structure, like in Table 4.5. The Ellipsis factor varied whether the critical region followed a pre-ellipsis site numeral (in the +NPE conditions) or an overt noun in the post-numeral position (in the -NPE conditions), where the overt noun was always identical to the NPE antecedent in the +NPE conditions. The critical region was always a preposition following the ellipsis site or corresponding overt noun, like in Experiment 3; spillover regions constituted the two words following this critical preposition. A sample item set is given in Table 4.5.

All conditions involved identical focus structure, with syntactically parallel objects contrasting a PP-internal numeral (*the inventor with FIFty gizmos*) in the first coordinate with a PP-internal numeral in the second coordinate (*the one with THREE gizmos*). We used *one*-anaphora within the second coordinate to discourage accent assignment prior to the right edge of the coordinate, per Harris and Carlson's (2016) observation that phrase-final contrast is strongly preferred (see §4.1.2.4 for further discussion).

The +NPE conditions always required resolving the antecedent of the ellipsis site to the PP-internal noun in the first coordinate (*gizmos*). As discussed above, there are multiple incremental processes operative in these stimuli: (i) resolving focus-sensitive ellipsis (if this occurs incrementally), (ii) building the focus structure of each coordinate and establishing contrast between them, (iii) predicting and assigning the focal accent to the correct position (the numeral), and (iv) resolving noun phrase ellipsis, which is a process that is dependent on (ii) and (iii). Because of this, I lay out assumptions about incremental focus structure and accent assignment shortly, and discuss how these processes may interact with our critical predictions.

Experimental items were counterbalanced across 4 Latin-squared lists and presented along with 54 fillers, for a total of 102 trials per experimental session. Fillers contained other constructions with medial prosodic boundaries (cued by commas), focus-sensitive coordination, and focus-sensitive operators (like *only*). 1/4 of the fillers involved appositive and restrictive relative clauses, some with ellipsis, like in Experiments 2-3. Some of these fillers contained noun phrase ellipsis, and others involved anaphora following a numeral (e.g., *two of them*).

4.1.2.3 Procedure

Participants read sentences in the Maze task, like in Experiments 2-3, and answered comprehension questions probing different regions of each sentence. Each sentence was followed by a comprehension question.

4.1.2.4 Predictions

The Context-Sensitive Retrieval Hypothesis predicts an interaction such that ellipsis resolution in NOBA (which separates its coordinates with a prosodic boundary in addition to a focus sensitive coordinator) should be easier than ellipsis resolution in

AA, where the lack of prosodic boundary between coordinates should be less likely to encourage partitioning of each coordinate into a distinct context in memory, if prosodic segmentation alone is a strong enough cue to trigger a shift in encoding context. Thus, CSR predicts faster decision times on the critical/spillover regions for the +NPE-NOBA condition relative to the +NPE-AA condition.

The Visibility hypothesis should make the opposite prediction: that the lack of prosodic separation of coordinates in *and also* should facilitate ellipsis resolution in the +NPE-AA condition relative to the +NPE-NOBA condition, because NOBA's prosodic boundary should render content from the first coordinate relatively less accessible.

Under both hypotheses, we expected a main effect of ellipsis; that is, the +NPE conditions should result in slower decision times at the critical region than the -NPE conditions, following the pattern of results observed in Experiment 3.

As detailed above, we assumed that there were several other linguistic processes operative in focus-sensitive coordination/ellipsis processing, including the incremental assignment of focus structure and focal accent position. Based on previous work on the processing of focus-sensitive coordination, we are in a decent position to make some baseline assumptions about how the processing of such constructions proceeds incrementally. Recall that in these particular experimental items, the final focus structure for both constructions should be identical, as in (26).

(26) a. Imala trusted not only [the inventor with FIFty gizmos]$_F$, but also [the one with THREE <~~gizmos~~>]$_F$ after the collaboration last winter.

b. Imala trusted [the inventor with FIFty gizmos]$_F$ and also [the one with THREE <~~gizmos~~>]$_F$ after the collaboration last winter.

Like in Experiments 2-3, Experiment 4 uses noun phrase ellipsis (NPE) to probe cross-boundary dependency resolution. We opted to use NPE for this experiment for two reasons: (i) again, because the size of the antecedent (a noun phrase) is small enough to be contained within the correlate of FSC constructions, and (ii) to ensure maximal comparability across the results of Experiments 2-4.

Recall that the default preference in focus-sensitive coordination constructions is for the contrastive foci (and associated pitch accents) to be in the final position of each coordinate (Harris & Carlson, 2016). Therefore, we assume that the least costly focus structure should contain a focal pitch accent on the final content word of each phrase. Furthermore, we expected that incremental accent assignment should matter for the implicit prosodic representation (Breen & Clifton, 2011; Van Handel, 2022). In our items, we attempted to ensure that the position of the focal accent was as close as possible to the right edge of each coordinate. Like Carlson and Harris's (2018) items, where the presence of *one*-anaphora required a pre-right edge accent on the preceding adjective, the presence of an NPE site at the right phrasal edge of our items required accenting a preceding numeral, as elided content cannot be accented (Rooth, 1992a). Because our critical region was on the immediately following word, it's possible that effects observed on this region may in part be due to reanalysis of the focal accent position within coordinate one. That is, if accent is incrementally assigned on content words, decision times on the critical region may include the cost of shifting the focal accent leftward to the licensing numeral, and performing this same operation in the phrase containing the antecedent. Because NPE cannot target a constituent that was previously encoded with a focal accent, we assume that resolving the focus structure of both coordinates must precede the resolution of NPE. Therefore, we propose the incremental timecourse for the processing of NOBA construction in Table 4.4.

Suppose that focal pitch accent within each coordinate is incrementally shifted along with each incoming content word until the parser reaches the end of the phrase. At time t_9 in Table 4.4, the accent should therefore be on *gizmos*. Nothing special should happen at $t_{12} - t_{14}$ in terms of accent placement, as these are function words

t_1 t_2 t_3 t_4 t_5 t_6 t_7 t_8 t_9 t_{10} t_{11}

Imala trusted not only [the inventor with fifty **GIZMOS**]$_F$, but also...

Word	+NPE-NOBA	-NPE-NOBA
t_{12}: [*the*	–	–
t_{13}: *one*	retrieve *inventor*	retrieve *inventor*
t_{14}: *with*	–	–
t_{15}: *three*	**THREE**	**THREE**
t_{16}: __ / *gizmos*	(ellipsis site)	! **GIZMOS** shift accent: $t_9 \rightarrow t_8$
t_{17}: *after* (crit.)	! shift accent: $t_9 \rightarrow t_8$ retrieve *gizmos*	–
t_{18}: *the* (spill.)	–	–

Table 4.4: Word-by-word timecourse of incremental focal accent assignment for NOBA conditions. ***Key:*** – = no initial accent; **bold** = (predicted) focal accent

that should indicate that the parser has not yet reached the contrastive focus within the second coordinate. At t_{13}, the antecedent of the anaphor *one* should be retrieved (*inventor*). At t_{15}, the numeral *three* should receive an accent, because this is the first element that contrasts with the content of the first coordinate (*fifty*). If the default preference is for the right edge of each coordinate to bear focal accent, the parser should expect to initially assign accent to *gizmos*. This should result in a disconfirmed expectation at t_{16} in the -NPE condition (see the righthand column of Table 4.4), because *gizmos* cannot be accented in both the first and second coordinates. Comparison of the current coordinate with the structure and content of the first coordinate should prompt reanalysis of the focal accent position in the first coordinate, such that the initially assigned accent on t_9 is shifted to the previous numeral, at t_8. Parsing of the remainder of the -NPE condition should then proceed without issue.

In the +NPE condition, however, t_{17} provides the first available cue to the presence of an NPE site, which should trigger a search for an antecedent. Let's assume for the sake of simplicity that the cue-based retrieval mechanism uses the feature +PLURAL as a proxy cue to activate appropriate candidates for the antecedent, because something

FSE		~~Imala trusted~~ ($\diamond_1$)
ANA	inventor	one
NPE	gizmos	~~gizmos~~ ($\diamond_2$)
COR	the inventor with fifty gizmos	the one with three ~~gizmos~~
...not only the inventor with fifty gizmos, but also $\diamond_1$ the one with three $\diamond_2$ after...		

Table 4.5: Dependencies in Experiment 4 stimuli. ***Key***: FSE: focus-sensitive ellipsis; ANA: anaphora; NPE: noun phrase ellipsis; COR: correlate; ~~elided content~~.

plural must be integrated into the ellipsis site in the post-numeral position (see also Kroll (2020) for evidence that plural NPE antecedents are preferred). As suggested for Experiments 2-3, syntactic parallelism across the coordinates may guide this process as well. At this point, several dependent processes should be triggered: (i) reanalysis of accent position in the first coordinate (deaccenting of *gizmos* and accenting of *fifty*), (ii) retrieval of the NPE antecedent (*gizmos*), and (iii) the search for a correlate (*the inventor with fifty gizmos*). Complete NPE resolution depends on process (i), because ellipsis cannot target focused material, and retrieval of the correlate depends on processes (i) and (ii), under the assumption that the structure of the remnant guides the search for the correlate. It's not entirely clear when the resolution of focus-sensitive ellipsis takes place, but if it occurs at the earliest possible point (immediately following the coordinator, *but/and also*), we assume that this precedes the critical region in the current structures. Overall, then, this results in the dependencies depicted in Table 4.5, in the following order: (i) focus-sensitive ellipsis resolution, (ii) *one*-anaphora resolution, (iii) NPE resolution, and (iv) remnant-correlate matching. In addition, we assume that accentual reanalysis precedes NPE resolution.

Though Table 4.4 only includes the NOBA conditions, we assume identical processes for resolving the focus structure and accent position once the parser reaches the second coordinate of *and also*-structures, at which point it should be clear that the sentence involves focus-sensitive coordination. Note then that any effect of focus structure or accentual reanalysis at or near the critical region should be constant across both structures, although the timecourse may vary slightly across +/-NPE conditions (note that accentual reanalysis is predicted to occur one word earlier in the -NPE condition; see Table 4.5).

Alternatively, focal accent reanalysis may be more costly for NOBA than AA, because NOBA contains a focus-sensitive operator in the first coordinate, whereas AA does not. This may encourage deeper and earlier commitment to the first coordinate's focus structure in NOBA compared to AA, which would predict a larger accentual reanalysis cost at the pre-critical region (*gizmos*) in -NPE conditions for NOBA compared to AA. If this pre-critical pattern obtains in the -NPE conditions, then one of two patterns may be borne out at the critical region in the +NPE conditions. Under CSR, facilitation of access to segmented material should benefit NPE resolution, whereas the cost of focus structure reanalysis should hinder NPE resolution. In principle, these effects could cancel each other out, resulting in no difference between NOBA vs. AA in the +NPE conditions. Under Visibility, both the presence of an intervening boundary and the cost of focus structure reanalysis should be a detriment to NPE resolution. Crucially, though, I predict that if such a difference between NOBA vs. AA structures holds, this should be observable in both the +/-NPE conditions, because the need to reanalyze focus structure should be present even in the absence of NPE. I return to a more thorough discussion of these points in the context of the experimental results.

4.1.3 Results

Average comprehension question accuracy was 81% (85% for the experimental items and 77% for the fillers).

Results are plotted in Figure 4.1. Bayesian linear mixed effects models[27] using

[27] `brm(logRT ~ Structure*Ellipsis + (1 + Structure*Ellipsis | Subject) + (1 + Structure*Ellipsis | Item))`

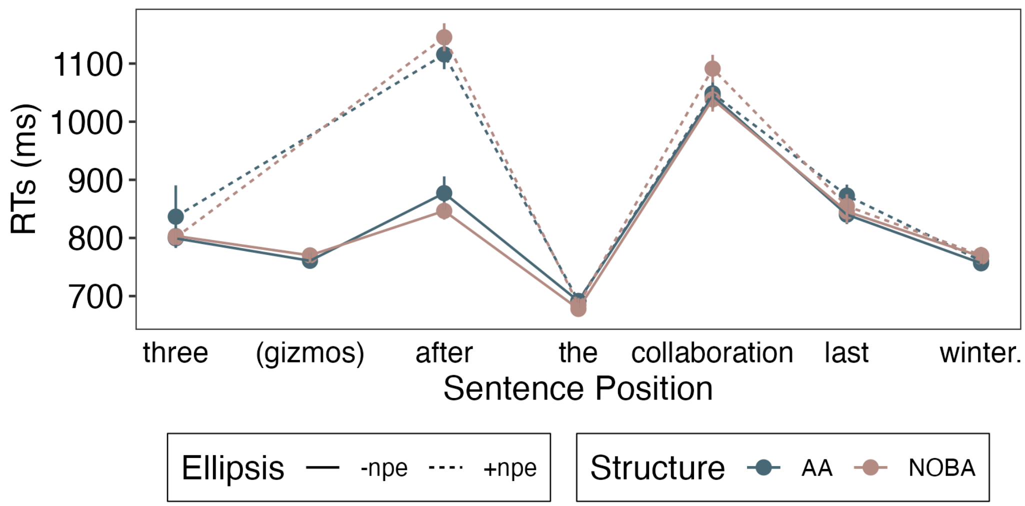


Figure 4.1: Word-by-word Maze latencies by condition for Experiment 4.

`brms` (Bürkner, 2017) were fit to response latencies at the critical region and the two spillover regions using the maximal random effects structure (Barr et al., 2013). Predictors were sum-coded, with -NPE and NOBA conditions coded as negative. Table 4.6 contains posterior estimates of the fixed effects along with 95% credible intervals. The $\hat{R}$-diagnostic and posterior predictive checks indicated that the models converged (Gelman et al., 2014).

Model results indicate a credible main effect of NPE at the critical region ($\hat{\beta}$ = 0.25, CrI = [0.2, 0.3]), but no effect of Structure ($\hat{\beta}$ = -0.01, CrI = [-0.02, 0.05]) and no Structure x NPE interaction ($\hat{\beta}$ = 0.04, CrI = [-0.04, 0.11]). At the first spillover region, no credible effects emerged. In the second spillover region, we observe an emerging interaction between Structure x Match ($\hat{\beta}$ = 0.05, CrI = [-0.01, 0.11]) without a main effect of NPE. Here, response latencies to the +NPE-NOBA condition were longer than the +NPE-AA condition.

Visual inspection of the pre-critical region suggested no difference between NOBA vs. AA structures, contra the focus structural reanalysis prediction discussed in the previous section. Because the second spillover region displayed a numerical trend

Effect	Critical (*after*) $\hat{\beta}$	95% CrI	Spill1 (*the*) $\hat{\beta}$	95% CrI	Spill2 (*collaboration*) $\hat{\beta}$	95% CrI
NPE	**0.25**	**(0.2,0.3)**	0.002	(-0.02,0.03)	0.01	(-0.02,0.04)
Structure	-0.01	(-0.02,0.05)	-0.01	(-0.04,0.01)	0.02	(-0.02,0.05)
NPE*Structure	0.04	(-0.04,0.11)	-0.003	(-0.05,0.05)	0.05	(-0.01,0.11)

Table 4.6: Bayesian linear mixed-effects models fit to log RTs at the critical and spillover regions of Experiment 4.

consistent with a late penalty for focus structure revision in +NPE conditions, we conducted a Bayes Factor analysis to quantify the strength of evidence in favor of an interaction at this region. This analysis compared a model with the interaction term to one without the interaction term, following the procedure specified by Wagenmakers et al. (2018) using the `bridgesampling` package in R (Gronau, Singmann, & Wagenmakers, 2017). This analysis resulted in a Bayes Factor (BF_{10} = 0.09) that indicated strong evidence for the non-interaction model (Lee & Wagenmakers, 2014). Overall, we take this to suggest that the results do not reliably reflect costs associated with focus structure revision at the critical or spillover regions.

Prior to the critical region, we observed inflated reading times in NOBA conditions at the anaphor *one* (see Figure 4.2). A post-hoc analysis (Table 4.7) fit a `brms` model[28] to this region, using treatment coding for the Structure predictor, with AA as the reference level. This analysis revealed a credible difference between structures such that the cost of anaphora resolution was greater for NOBA than for AA. This effect is notable, as it potentially provides evidence for Visibility or a focus structure revision cost; the discussion further elaborates on these possibilities.

[28]`brm(logRT ~ Structure + (1 + Structure | Subject) + (1 + Structure | Item))`

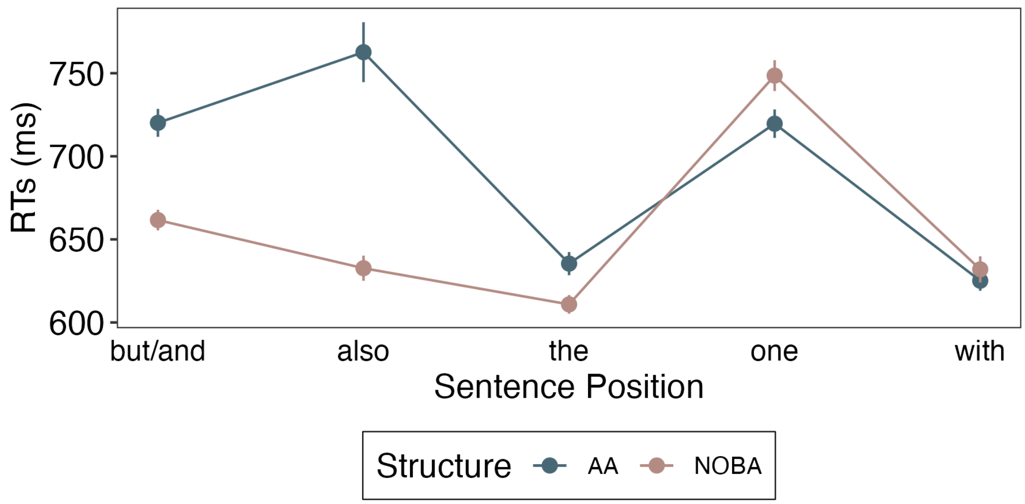


Figure 4.2: Pre-critical word-by-word Maze latencies by Structure for Experiment 4, on regions containing the focus-sensitive coordinators and *one*-anaphor.

	Pre-critical (*one*)	
Effect	$\hat{\beta}$	95% CrI
Structure	**0.04**	**(0.01,0.07)**

Table 4.7: Post-hoc Bayesian linear mixed-effects models fit to log RTs at the pre-critical *one*-anaphor region of Experiment 4.

4.1.4 Discussion

Recall the predictions of the CSR hypothesis: NPE resolution in NOBA should be faster than NPE resolution in AA, if the presence of a prosodic boundary triggers a shift in encoding context, and thus reduces contextual interference immediately following retrieval of an antecedent. The results of Experiment 4 do not support CSR, as we observe no difference between NPE resolution in NOBA vs. AA at the critical region. Similarly, the predictions of Visibility were not borne out at the critical region. Under this hypothesis, we should have observed longer NPE resolution times for NOBA due to prosodic separation of the antecedent and ellipsis site. We also did not observe any evidence of focus structure revision at the pre-critical region in the -NPE conditions or at the critical/spillover regions in the +NPE conditions. Recall that although there was a numerical penalty for +NPE-NOBA at the second spillover region, a Bayes Factor analysis did not provide support for the validity of this trend. We did, however, find evidence of a Visibility-like effect at the pre-critical anaphor *one*, where latencies to NOBA structures were longer. I return to a discussion of this shortly.

As was argued in the discussion of Experiment 3, it's possible that even if prosodic boundaries reliably trigger context shifts, the effects of contextual interference per CSR may not be observable during sentence processing, because item-level cues outweigh contextual cues. If there is a small but true effect of CSR for sentence-final ARCs, this once again leaves us in a position where we would have to account for a difference between ARCs and other segmented structures. One might wonder whether such a difference could be due to the fact that ARCs align with prosodic and discourse boundaries; we think this unlikely, because NOBA, too, involves coordination of proposition-sized units that make up coordinating segments in the discourse representation, linked via a CONTRAST relation. Alternatively, it may be that the ability to observe a mechanism like CSR at work may jointly depend on the global structure involved and the requirements of the particular dependency at hand. Recall that CSE proposes that prosodic boundaries serve to structure syntactic content in memory. We assume that NPE resolution, in part, requires access to a syntactic representation (N. Kim, Brehm, & Yoshida, 2019; C. Miller, 2016), but it is also a discourse-sensitive dependency (Kroll, 2020). Therefore, it may not offer the most optimal test case for probing CSR. As such, it would be useful for follow-up work to probe syntactic dependencies; Experiments 5-8 address this possibility.

I entertain another reason why we might have failed to obtain the predicted interaction at the critical region, namely the process of assigning implicit prosodic strucutre in the Maze task. Previous work has found that effects of implicit prosody replicate well in the Maze, but crucially, this was for syntactic attachment decisions based on later-stage prosodic parses (Van Handel, 2022). Recall that the Maze task encourages deeper incremental processing, and thus typically results in longer decision times per word than other reading tasks, like self-paced reading (Witzel et al., 2012). The fact that relatively slow, incremental decisions must be made at each word may encourage an abnormal first-pass prosody, where readers are encouraged to insert prosodic breaks early and often (see Van Handel, 2022, for more extensive discussion for and against this position).

Recall the discussion of potential prosodic structures for NOBA and AA from §4.1.1. Due to the nature of the task and the length of each coordinate in the experimental stimuli, it could be that participants were encouraged to insert a prosodic break before the second coordinate of AA in the final prosodic parse, collapsing the prosodic boundary difference between NOBA and AA. That is, readers may have inserted an implicit boundary in the pre-second coordinate position in both constructions. A replication of the current experiment in self-paced reading or eye-tracking would be beneficial for confirming that the lack of interaction at the critical region was not simply due to a task effect driven by the fact that the Maze may encourage an unnatural first-pass prosody. I leave this possibility for future work, but contend that there is a more likely reason for the lack of interaction at the critical region.

I return to the observation that we found inflated RTs on the anaphor *one* in NOBA conditions compared to our AA conditions. This is plotted in Figure 4.2, which also shows evidence of inflated RTs on the coordinator region for *and also* compared to *but also*. We attribute this to the fact that *but also* is more predictable in the context of the preceding focus particle *not only* than *and also* is. As discussed above, reaction times on *one* revealed a credible difference between structures such that the NOBA conditions exhibited longer RTs than AA ones. I entertain two possible explanations for this effect.

This difference could be due to an expectation for a focal accent on the highest position within the NP correlate, as in (27) which signals contrast between *inventor* and *engineer*, as opposed to the focus structure of our experimental items (*not only [the inventor with FIFty gizmos]$_F$*), where the numeral bore focal accent.

(27) not only [the invENtor with fifty gizmos]$_F$, but also [the engiNEER with fifty gizmos]$_F$

Such an expectation is possible but perhaps at odds with Harris and Carlson's (2016) suggestion that the preference is to establish contrast with a final element[29]. However, their structural configurations were slightly different than ours. They used ambiguous constructions that allowed for multiple potential simple NP correlates, but did not allow for multiple possible accent positions within each NP (28). In contrast, our constructions were unambiguous in the position of the correlate, but the internal structure of our NPs was more complex (29). In our stimuli, focal accent position

[29] Specifically, this preference applied to the search for a correlate. It may not extend to a preference for the position of focal accent.

was temporarily ambiguous (prior to the processing of the second coordinate) and potentially could have been positioned on NP1, the numeral, or the PP-internal noun.

(28) Harris and Carlson (2016) structural configurations

...$\mathrm{NP}_{\mathrm{CORR1}}$ [that V $\mathrm{NP}_{\mathrm{CORR2}}$], Coord...

(29) Experiment 4 structural configurations

...NP1 [$_{\mathrm{PP}}$P Num NP2], Coord...

Thus, the cost at *one* for NOBA could reflect a disconfirmed expectation for focal accent on NP1 in (29). Under this view, focal accent assignment must proceed differently in the incremental time course of *and also*, although the final focus structure for NOBA and AA is argued to be the same (Hulsey, 2008). Upon parsing the coordinator in AA, the parser must hold off on generating an expectation for accent in the highest NP of coordinate two, despite the fact that focus-sensitivity of coordinate two should be apparent at this point. Otherwise, we would expect the same penalty to apply in both AA and NOBA. Because the focus structure of coordinate one in AA may not be pre-determined in the same way as coordinate one of NOBA, it is possible that focal accent assignment for coordinate two is delayed until the full NP has been processed. If that were true though, we should expect a later cost to emerge for AA (e.g., at the right edge of the second coordinate), reflecting updating of the focus strucutre. This does not align with the decision time pattern we observe in our results. In fact, the inflated reading times at the coordinator region for *and also* could partially reflect the cost of updating of the focus structure of coordinate one, given that the coordinator provides the first cue to focus-sensitivity in this condition. In this case, we would expect that focus structure assignment for coordinate two of NOBA and AA proceeds in the same manner.

An alternative explanation is that the NOBA penalty at *one* reflects a Visibility effect, where retrieval of the antecedent *the inventor* was hindered only in NOBA, because it was encoded in a distinct context and was thus more difficult to access. This explanation, too, has its issues. Recall that Carlson et al. (2009) argued that Visibility effects apply only to syntactic parsing decisions, but not semantic dependencies (see §2.2.2 for an extensive discussion). A Visibility effect for an anaphoric dependency like *one*-anaphora would contrast with this view. However, Carlson et al. (2009) may have failed to find evidence of Visibility because their critical dependency involved the search for a correlate, a multi-stage process, rather than the retrieval of an item from memory. The number of linguistic processes active at the time of their sentence-final judgments may have obscured an underlying Visibility effect. Dependency-related concerns aside, under CSE-Visibility, we should expect that Visibility applies to any retrospective dependency resolution process. Furthermore, CSE-Visibility predicts only a distinction between the current encoding context and all previous contexts, whereas all of Carlson et al.'s stimuli contained at least one preceding prosodic boundary. Therefore, the boundary effect could have been eliminated in their experiment because all conditions required accessing a previous encoding context. CSE-Visibility then offers a way to explain the discrepancy between the current study and Carlson et al.'s results.

A Visibility-based explanation of the current results introduces a more pressing issue: it once more leaves us in a position where there is a mysterious difference between ARCs and other at-issue, segmented structures. Notably, retrieval across sentence-final ARC boundaries does not evidence Visibility, as evidenced by Experiments 2-3. It remains mysterious why a Visibility effect should arise for NOBA, which from a purely structural perspective, has boundary positions that are analogous to sentence-final ARCs. As discussed earlier, CSE predicts that they should result in identical contextual partitions. Schematic representations are repeated in (30).

(30) a. The author that wrote **two novels** met the hack, who wrote **forty** _

b. not only the inventor with **fifty gizmos**, but also the one with **three** _

The current results cannot distinguish between the two possibilities I've outlined here, but I suggest a potential follow-up design to tease them apart below. Setting aside the particular source of the NOBA penalty at *one* for the time being, I offer a related comment on this effect. A fundamental premise of the Temporal Context Model is that any item-level retrieval process involves reinstatement of that item's contextual features as well (Howard & Kahana, 2002). Then by hypothesis, contextual features associated with the first coordinate should be reinstated during the search for *the inventor*, the antecedent of *one*, in all conditions. The experimental stimuli always included *one*-anaphora in the NP1 position of the second coordinate in order to discourage positing an early focal accent prior to the right edge of each coordinate. However, this design choice may have inadvertently rendered any boundary effect at the critical NPE site inert, because the relevant contextual features of the target encoding context would have been pre-activated at *one*, per CSR. To rule out this possibility, the experiment should be replicated without an anaphoric dependency prior to the NPE site, like in (31).

(31) a. Imala trusted not only the inventor with ten gizmos, but also the inventor with fifty <~~gizmos~~> after the collaboration...

b. Imala trusted the inventor with ten gizmos and also the inventor with fifty <~~gizmos~~> after the collaboration...

This comparison could also serve to delineate between the possibilities suggested above: if the penalty at *one* has to do with disconfirmed focus structural expectations, the same penalty for NOBA should persist on *inventor* in the contrast above. In this case, if reinstatement of the previous encoding context at *one* was obscuring an effect of CSR at the critical region in the current experiment, facilitation of NPE in NOBA structures should emerge in the absence of the *one*-anaphor. If instead the penalty at *one* reflects a true Visibility effect, there should be no difference between (31a) and (31b) at *inventor*, but a penalty for the +NPE-NOBA condition should emerge following the ellipsis site.

The current study is not equipped to definitively determine why the boundary-crossing profile of sentence-final ARCs vs. NOBA may differ. I postpone a full discussion of potential cross-construction differences until the General Discussion (§4.4).

4.2 Experiment 5

Experiment 4 failed to find evidence that CSR extends to at-issue, prosodically segmented constructions. Recall that the premise of Visibility originally related to the accessibility of potential syntactic attachment sites, and suppose we extend this hypothesis to apply to cases where prosodic phrases group syntactic content in memory more generally. We might then expect that NPE (the dependency used in Experiments 2-4) is not the most appropriate dependency to probe the hypotheses detailed in §2.3.3, because it may not be a purely syntactic dependency. As such, I turn to the issue of how prosodic and discourse boundaries may constrain the resolution of syntactic dependencies, using coordinate agreement in another type of focus-sensitive coordinate (FSC) structure as a test case. The current experiment turns to an investigation of subject-verb agreement, comparing the processing of coordinate agreement in integrated coordinate structures like (33) to that of prosodically isolated *as well as* (AWA) coordination, like in (32).

(32) Ramona thinks that the chefs, as well as the butchers, have prepared an inno-

vative menu for the restaurant.

(33) Ramona thinks that the chefs and the butchers have prepared an innovative menu for the restaurant.

The remainder of this section briefly discusses the properties of focus-sensitive AWA constructions (§4.2.1), then outlines some relevant facts about the processing of ordinary coordination, which displays a preference for agreement with the closest conjunct (§4.2.2). If AWA coordinates display bypassing, like ARCs, we should expect sensitivity to features of the closest conjunct to be diminished for (32) compared to (33).

4.2.1 *...as well as...*

I assume that *as well as* (AWA) constructions (i) are focus-sensitive coordinate structures (De Vries, 2005; Hulsey, 2008; Krifka, 1998), and (ii) may behave like parentheticals[30] in some ways, but differ in important ways from other discourse asides, like ARCs and *and*-parentheticals. Therefore, they offer an example of prosodically isolated, sentence-medial phrases that are linked by a coordinating discourse relation (e.g., CONTRAST or PARALLEL), like *not only...but also.* I discuss each of these features in turn.

AWA-coordination, like other focus-sensitive coordinate (FSC) structures, involves coordination of propositions, focus-driven movement of a correlate-remnant pair, and ellipsis of remaining content within each coordinate (Hulsey, 2008). This is evidenced by the ungrammaticality of (34b) in contrast to (34a).

(34) a. I had to take the students, as well as their parents, to the auditorium.

[30]This term is underspecified, but here, I mean that they sometimes display syntactic, prosodic, or discourse independence.

b. *I had to take the students, as well as I had to take their parents, to the auditorium.

Thus, their syntactic analysis is argued by some to be identical to that of *not only...but also* and other types of focus-sensitive coordination (see the discussion of NOBA in §4.1.1 for more detail; the same syntactic diagnostics discussed there apply to AWA-coordination as well). This results in the structure in (35).

(35) [The CHEF]$_{F,1}$ $<t_1\ t_3>$, as well as [the BUtcher]$_{F,2}$ $<t_2\ t_3>$, [have prepared an innovative menu]$_3$.

The meaning of *as well as* is argued to involve ordinary coordination plus an additive component analogous to *also*, which presupposes the existence of at least one focus alternative (Krifka, 1998). Thus, these constructions have a meaning like *and also*, which was also introduced in §4.1.1. This is exemplified by (36), where co-occurrence of *as well as* and *also* leads to unacceptability (Hulsey, 2008). Like *and also* (37b), AWA lacks the collective reading possible under ordinary coordination, as exemplified by the contrast between (37a) and (37c).

(36) #Finnegan married Margot as well as also Poppy.

(37) a. Finnegan married Margot and Poppy.
Collective reading: one wedding (married to each other)
Distributive reading: two weddings

b. Finnegan married Margot and also Poppy. $\rightarrow$ collective unavailable

c. Finnegan married Margot as well as Poppy. $\rightarrow$ collective unavailable

AWA coordinates sometimes appear to pattern like other parentheticals, as the content of the second coordinate may be treated as syntactically independent. Like *and*-parentheticals (Kluck, Ott, & De Vries, 2014), when isolated AWA-coordination occurs in subject position, the verb can but need not enter into an agreement relation with the content of the second coordinate. That is, number features on the verb and N2 need not match. Consider the optionality in plural-marking on the verb in (38a) and (38b), but not (38c). That is, singular marking on the verb is permitted with a coordinate subject for both *and*-parentheticals and *as well as*-coordination, but not for ordinary coordination.

(38) a. The chef – and the butcher, too – {has, have} prepared an innovative menu.

b. The chef, as well as the butcher, {has, have} prepared an innovative menu.

c. The chef and the butcher {*has, have} prepared an innovative menu.

Some caveats apply here. First, the presence of plural number features on the second coordinate seems to affect agreement (39), suggesting that the verb is not totally insensitive to the content of N2. Second, it's less clear that N2 features are less relevant for integrated AWA-coordination (40), but to my ear, this at least sounds more acceptable than (38c).

(39) ??The chef, as well as the butchers, has prepared an innovative menu.

(40) ?The chef as well as the butcher has prepared an innovative menu.

If these structures involve ellipsis, then the fact that singular number marking on the verb is permitted in (38b) could be captured by a structural analysis like in (41). However, the presence of ellipsis is not compatible with plural number marking (42), suggesting that AWA coordination may be structurally ambiguous in whether it involves ellipsis or not.

(41) The chef <has prepared an innovative menu>, as well as the butcher <has prepared an innovative menu>, has prepared an innovative menu.

(42) *The chef <have prepared an innovative menu>, as well as the butcher <have prepared an innovative menu>, have prepared an innovative menu.

Together, we take these facts to suggest that the second coordinate is involved in subject-verb agreement.

In addition, AWA-coordinates often occupy a distinct prosodic phrase, whereas N2 of ordinary coordination does not. Though AWA-coordinates may be prosodically separated – via commas, parentheses, or em dashes in written text – this is not a requirement (see (43) below). It's possible that the presence of commas affects the information structure of an AWA-containing sentence, namely whether the second coordinate is interpreted as central to the main point of the sentence or not. Because no literature to my knowledge explicitly discusses these issues, and because they are central to the design and results of the study reported in this section, I provide a preliminary discussion here.

Like in the case of ARCs, it intuitively seems that at least for sentence-medial cases, AWA-coordination can be accompanied by pauses, boundary tones, pitch compression, or otherwise distinct intonation. Also like ARCs, it is likely that there is no one intonation that is uniquely assigned to AWA-coordinates. It's not clear that prosodically integrated cases of AWA-coordination don't also involve some degree of prosodic separation (e.g., φ-phrase boundaries and accompanying phrasal tones at the right edge of each coordinate). As has been argued for other clausal/propositional units, I assume that (i) AWA-coordinates, when prosodically isolated, are bounded by ι-phrase boundaries and that (ii) those instances of AWA-coordination that are not prosodically set off in speech or comma-marked in writing contain weaker prosodic boundaries (i.e., φ phrases) than prosodically isolated instances. As such, the present experimental stimuli were constructed under the assumption that the presence of commas in sentences with AWA-coordination corresponds to distinct treatment in their implicit prosodic representation. Recall that distinct implicit prosody for appositives is corroborated by Hirotani et al. (2006) in an eye-tracking study; the self-paced reading results of Experiment 6 show that isolated AWA-coordination displays the same reading time pattern. Nevertheless, future listening and production studies should be conducted in order to confirm the prosodic properties of AWA-coordination in production and the correspondence between their overt and implicit prosody.

As was argued for *and also* constructions, it's likely that the size of each coordinate for non-comma-marked AWA-coordination contributes to the likelihood of a prosodic break in the pre-coordinate two position. The examples in (43)-(46) suggest that coordinate length and intended discourse function (broadly construed) may both play a role in the presence of commas in the orthography of sentences with AWA-coordination, but it doesn't seem that AWA-coordinates are interpreted as asides in the way that ARCs are. The evidence to follow points to the conclusion that AWA-coordinates are not discourse-subordinate segments. Even when they seem to contain secondary information in an intuitive sense, they clearly contribute contrastive content, as is standard for additive focus particles (Krifka, 1998). This suggests that they relate to a common higher-level QuD, like in contrastive focus/topic structures (Büring, 2003).

(43) "The forces transmitted via the sails are resisted by forces from the hull, keel, and rudder of a sailing craft…This combination of forces means that it is possible to sail **an upwind course as well as downwind**."
(Source: *Sailing*. From Wikipedia.)

(44) "Grant similarly works with **children and teens, as well as their parents**, on healthy digital device management."
(Source: *Media Overload is Hurting our Mental Health*)

(45) "Given **the encouraging results from this first test, as well as the qualitative research that validated our approach**, we decided to roll the feature out to news subscribers across our platforms."
(Source: *A New Way to Share New York Times Stories*)

(46) **Q**: "Why are people so obsessed with cat behavior? They're going to be unreliable in their affections. Unlike dogs who are friends for life. Dogs are selfless; cats are selfish."

> **A**: "You are so wrong!...**Cat love (as well as dog love)** is unconditional. **With a cat**, though, you might have to show that you are worth such love before it is given." (Source: *Why are people so obsessed with cat behavior?* From Quora.)

The content within the second coordinate of (43), a non-comma-marked example, contains only one overt word (*downwind*). This coordinate structure conveys the meaning *an upwind course and also a downwind course*, without "backgrounding" either coordinate. Comma-marked AWA constructions with only one content word are also attested, as in (44). Note in (44) that the content of the second coordinate may be thought of as secondary in some sense, as the main topic of the article is the effect of headline stress on teens and young adults (as evidenced by the opening sentence, *Many of the teens and young adults [Grant] has worked with...*), not their parents. Example (44) is the first mention of working with children in the article. However, I assume that the main QuD of the sentence is *Who does Grant work with on healthy digital device management?* Therefore, while children and teens are more central to the discourse topic of the article, the QuD corresponding to coordinate two corresponds to *Who else does Grant work with?* This suggests that the coordinates *children and teens* and *their parents* are meant to be interpreted contrastively (i.e., they contribute two answers to the same QuD), and perhaps readers draw the additional inference that working with children and teens also requires working with their parents.

In contrast, (45) contains prosodically isolated coordinates that jointly contributed to the decision referenced in the subsequent clause. However, the amount of content contained within each of the coordinates in this example is greater than that of examples (43) and (44), which may have contributed to the decision to phrase the coordinates separately (evidenced by the presence of commas).

Finally, the example in (46) contains AWA-coordination that is set off by parentheses in the orthography. Like (44), the example in (46) appears to contribute a secondary point in the sense that the sentence contrasts cats with dogs, but the the writer continues on with cats as the topic of the following sentence (*With a cat, though...*). This seems to serve the purpose of agreeing with the questioner on their point about dogs (that they are selfless), but disagreeing on the point about cats (that they are selfish). Again, this points to a contrastive relationship between the first and second coordinates.

Similarly, in the response to (47), the *as well as*-coordinate does not directly respond to the preceding QuD.

(47) Q: What did the chefs do?

A: The chefs – as well as the butchers, by the way – prepared an innovative menu for the restaurant.

Instead, it likely answers a PARALLEL question, typical of contrastive focus-based relations (Brunetti, 2024), like: *Who else prepared an innovative menu?* Crucially, this QuD is focus-congruent with the main QuD. This is a central property of AWA-coordination; it cannot contribute focus-incongruent content, unlike ARCs, which may address a completely irrelevant QuD, as in (48). As argued by Repp (2016), CONTRAST relations fundamentally require "similarities as well as dissimilarities" (p. 277). The fact that AWA-coordination requires a component of similarity between discourse segments aligns with the view that they involve focus-sensitive ellipsis, which is constrained by parallelism. Note that the AWA counterpart of (48) in (49) is ungrammatical.

(48) Noodle, who loves cheese, was taking a nap.

a. Main QuD: *What was Noodle doing?*

b. Secondary QuD: *What does Noodle love?*

(49) *Noodle was taking a nap, as well as loves cheese.

AWA-coordination diverges from the profile of ARCs in other ways as well. For a complex QuD like in (50), either the response in (50a) without prosodic demarcation of coordinate two, or the response in (50b) with prosodic demarcation seem to be equally acceptable. Contrast this with attempting to insert given information into an appositive relative clause, like in (51), which leads to infelicity. If prosodically isolated AWA coordinates are always interpreted as discourse asides, we should expect similar infelicity to arise in (50b), but this does not seem to be the case.

(50) Q: What did the chefs and the butchers do?

  a. A: The chefs as well as the butchers prepared an innovative menu for the restaurant.

  b. A: The chefs, as well as the butchers, prepared an innovative menu for the restaurant.

(51) Q: What did Gigi and her sous chef do?

A: #Gigi, who has a sous chef, prepared an innovative menu for the restaurant.

There are also notable differences in the distribution of AWA-coordination and *and*-parentheticals from a focus structural and discourse perspective[31]. This is exemplified by the following examples from Blakemore (2005), modified to include AWA-coordination in the (b) counterparts.

(52) a. I had to take the whole class – and I'm talking about a hundred students – all around campus until I found an empty lecture theater.

  b. *I had to take the whole class – as well as I'm talking about a hundred students – all around campus until I found an empty lecture theater.

(53) a. We were out in the garden and [a big rat]$_1$ – and I mean [a BIG rat]$_1$ – ran out from under the table.

[31]These examples differ from the *and*-parenthetical in (38a), which is argued to involve parenthetical-internal stripping (Kluck et al., 2014). The presence of ellipsis and the additive particle *too* suggests that this example is like *as well as*. But, *and*-parentheticals can establish contrast without focus-sensitive ellipsis (*I had to take the whole class – and I'm talking about their parents, too – all around campus*), whereas AWA cannot. Contrast this with (52b).

b. #We were out in the garden and [a big rat]$_1$ – as well as [a BIG rat]$_1$ – ran out from under the table.

Because AWA-coordinates are strongly constrained by syntactic parallelism and involve obligatory focus-sensitive ellipsis, (52b) is ungrammatical. Additionally, the unacceptability of (52b) and (53b) can be attributed to the fact that AWA requires that its coordinates stand in a contrastive relationship. In both cases, the content of coordinate two refers to the same entity as in coordinate one. This type of repetition for the sake of emphasis is not permitted in AWA-coordination. The examples above show that AWA-coordinates (i) are structurally distinct from *and*-parentheticals and (ii) serve a different pragmatic function than they do.

I take the combination of facts presented here to suggest that while the second coordinate of AWA can be prosodically set apart within a separate ι-phrase and contribute secondary information relative to the main topic of the discourse (in an atheoretical sense), AWA-coordination is not "parenthetical" in the sense of ARCs or *and*-parentheticals. Instead, I will assume that the coordinates of AWA are discourse segments linked via a PARALLEL relation, because they offer parallel answers to the same QuD. Therefore, I assume a discourse structure as in (54).

(54) QuD Structure for AWA-coordination in Experiment 5

Who prepared an innovative menu?

[The BUTchers]$_F$ prepared an innovative menu. [The CHEFS]$_F$ prepared an innovative menu.

While a thorough discussion of AWA-coordination is beyond the scope of this dissertation, it is clear that these constructions diverge in notable ways from other better-studied parentheticals. Future work should further investigate their syntactic, prosodic, and discourse properties in context.

4.2.2 Closest Conjunct Agreement

Before turning to the details of Experiment 5, this section discusses some properties of subject-verb agreement for ordinary coordinate structures, which serve as the baseline prosodically integrated conditions in the present experiment. One possible view of coordinate agreement within a cue-based retrieval framework is that the retrieval cues always target the highest NP. Under such a view, it's unlikely that the features of the nouns contained within a coordinate subject would be accessed independently, perhaps because they are re-encoded as a single unit at some higher level of representation. In line with this view, Garrod and Sanford (1982) show that processing a plural pronoun (*they*) following a conjoined subject (*John and Mary*) is easier than processing a singular pronoun (*he/she*), suggesting that there is a cost associated with accessing a single NP antecedent in these structures. This provided early evidence for access to a complex discourse representation for pronominal anaphora. But, this does not necessarily suggest that the retrieval cues recruited for subject-verb agreement also privilege the highest NP, as the processing of agreement is thought to involve retrieval of matching number features in English.

For instance, in structures such as [$_{\text{NP}}$ N1$_{\text{+PL}}$ and N2$_{\text{+PL}}$], if the retrieval probe on the following verb targets any plural NP, note that there are three potential candidates for retrieval: N1, N2, and the NP that dominates them. For the retrieval probe to target only the highest NP, the retrieval mechanism itself would have to have access to information about the higher-order syntactic structures previously encountered in the sentence (e.g., some notion of: this is a sentence with a complex subject), or perhaps, a feature like +HIGHEST[32] is always operative and weighted heavily, but is rendered vacuous in the case of simple subjects because the only available NP is also

[32] I propose this cue only for the sake of argumentation, and do not intend to make any commitments about its existence. This merely stands in as a proxy, as other possible syntactic cues like +NOMINATIVE or +SPEC-TP may not uniquely identify the highest NP in this case.

the highest one.

Although it is generally possible to target syntactic phrases above the word level (e.g., +VP) it cannot be the case that a retrieval cue like +HIGHEST is generally operative such that only the features of the highest syntactic category matters. Recall the case of agreement attraction, where the features of a PP or RC-internal noun may interfere with subject-verb agreement during retrieval, as in (55) (Wagers et al., 2009). This highlights the fact that the retrieval mechanism is error-prone, and may sometimes access items that are structurally illicit.

(55) The key [$_{\text{PP}}$ to the **cabinets**$_{\text{+PL}}$] **were**$_{\text{+PL}}$ rusty.

Two additional pieces of experimental evidence validate the idea that subject-verb agreement in coordinate structures is sensitive to lower-level syntactic features of subject-internal nouns, and may be prone to accessing one noun to the exclusion of the other in the early timecourse. First, Wagers and McElree (2022) find that even added material in DP-internal structure can cause earlier dependents to be shunted from the focus of attention, more or less immediately, necessitating retrieval later on. Though Wagers and McElree's study investigated intervening content between a determiner and noun, it's possible that the same holds for the coordinates of a complex subject. That is, that N1 and N2 of a complex subject are shunted from the focus of attention prior to encountering the verb, especially as distance between the second coordinate and the verb increases. If we assume that the retrieval cues on the verb are {+PLURAL, +NP}, for example, there is nothing to stop the subject-internal nouns from being independently accessed during retrieval. In the case of prosodically isolated coordinates, initial access to one coordinate may inhibit the parser's ability to access the other. Therefore, in *N1*$_{+PL}$*, as well as N2*$_{+SG}$*,...V*$_{+PL}$ structures, N2 may never be accessed, because the cue +PL does not serve as a good cue to the AWA-coordinate-internal noun, and accessing the first coordinate will not serve as a good contextual cue to the second one. Under such a model, some notion of the global syntactic structure (i.e., that N1/N2 are part of a complex subject) must be recruited in order to fully integrate the subject with the verb. The same principle may hold true for integrated coordinate structures as well, but there, the lack of prosodic separation should (by hypothesis) allow for easier access to both N1 and N2, even if only one is probabilistically accessed at an early stage, because they should be very similar in terms of contextual features. Experiments 7-8 further investigate the consequences of positing such a mechanism, especially for the processing of isolated AWA-coordination.

The second piece of evidence comes from work by Keung and Staub (2018) on coordinate agreement, which suggests that the syntactic features of each noun within a complex subject can affect the processing of agreement. Across a series of experiments, Keung and Staub investigate the contrasting intrusion profiles of standard NP-PP agreement attraction configurations with NP-and-NP coordinate structures in English, ultimately finding a preference for configurations where the closer conjunct matches the verb in (plural) number features compared to when it mismatches the verb, in both production and comprehension. This is exemplified in Table 4.8. They analogize this effect to cases of closest conjunct agreement (CCA) in languages where there is a grammatical pattern of agreement (of number, person, or gender features) with the linearly closer NP in coordinate structures (Nevins & Weisser, 2019). The CCA effect surfaced both in the continuations produced or selected by participants in sentence completion and 2AFC completion tasks, as well as in reading measures at the critical verb in an eye-tracking study. Furthermore, they find that distance between the second coordinate and the verb slightly diminishes, but does not eliminate, the effect, suggesting that the verb need not be proximal to the intervening noun for mismatching second coordinates to result in processing disruption at the verb.

For the purposes of the current discussion, I focus primarily on the reading time

Structure	Sentence	RT Pattern
PL-PL	The maids and the butlers are laboring...	PL-SG > PL-PL
PL-SG	The maids and **the butler** are laboring...	
	→ *grammatical disruption effect*	
*PL-PL	*The maids and **the butlers** is laboring...	*PL-SG < *PL-PL
*PL-SG	*The maids and the butler is laboring...	
	→ *ungrammatical facilitation effect*	

Table 4.8: Closest Conjunct Agreement pattern for coordinate structures from Keung and Staub (2018).

.

effects for coordination reported by Keung and Staub (2018) (K&S), as these served as a starting point for predictions in the present experiment. In first-pass and go-past times, the authors found that mismatching second coordinates resulted in slower RTs on grammatical verbs. That is, $N1_{PL}$-$N2_{SG}$ subjects resulted in slower reading times at a verb bearing a plural feature; following K&S, I term this the *grammatical disruption effect*. In total reading times, mismatching second coordinates resulted in faster RTs on ungrammatical verbs. In other words, ungrammatical conditions where the second coordinate matched the verb, as in *PL-SG in Table 4.8, resulted in faster reading times relative to when both nouns mismatched the verb, as in *PL-PL; I term this the *ungrammatical facilitation effect*. Furthermore, Keung and Staub (2018) show that this pattern of results contrasts with the standard pattern for agreement attraction, where there is typically a grammaticality asymmetry: namely, a locally matching intervener facilitates processing of an ungrammatical verb, but does not disrupt processing of a grammatical verb (i.e., *The key to the cabinets is...* does not result in slower RTs at the verb). In coordinate structures, however, both ungrammatical and grammatical agreement with the verb is affected by mismatches on the closer noun.

Why, then, might the computation of subject-verb agreement in coordinate structures be unique? K&S hypothesize that like languages that grammatically adhere to agreement with the closest conjunct, the English processor also adheres to an agree-closest preference. This aligns with Willer-Gold et al.'s (2018; 2016) analysis of CCA in South Slavic, where competition between grammatical principles for number-marking on the verb in coordinate structures – AGREE-CLOSEST, AGREE-HIGHEST (the grammatical counterpart of the hypothetical +HIGHEST cue discussed above), and default (plural) agreement – leads to slower production latencies as the degree of mismatch between the output of each principle increases. Thus, they argue that even for languages with grammatical CCA, processing factors interact with agreement. Alternatively, K&S suggest that perhaps the ultimate computation of number-agreement is determined semantically, in that notionally plural subjects receive default plural marking on the verb.

In a cue-based retrieval framework, we could think of the processor probabilistically retrieving either coordinate and subsequently reactivating and retrieving the complex subject as a whole, when both coordinates match in phi-features with the verb. When either of the coordinates mismatch, one possibility is that the one that matches the verb is retrieved with greater success, and by the same process, the rest of the complex subject, which also contains the mismatching noun, is subsequently reactivated and must be marked for semantic plurality before subject-verb agreement is fully resolved. Such a process would also explain why mismatching coordinates lead to processing disruption at the verb. Additionally, as suggested by Willer Gold et al. (2018) and K&S, it may be the case that the AGREE-CLOSEST principle is more strongly weighted than the competing principles for grammatical reasons.

We used the CCA effect as a starting point for the current study. We aimed to first replicate Keung and Staub's results for coordination using bidirectional self-paced reading, then determine whether encapsulating the second coordinate in a prosodically-separated *as well as* phrase would enable the parser to more easily "bypass" the mismatching second NP, in an analogous manner to the bypassing effects that have been shown for appositive relative clauses and nominal appositives. If so, this should lead to a reduction in the CCA pattern found by K&S for ordinary coordination: i.e., (i) in grammatical conditions, a diminished grammatical disruption effect for parentheticals relative to their integrated counterparts, and (ii) in ungrammatical conditions, a diminished ungrammatical facilitation effect for parentheticals relative to integrated coordinate structures.

4.2.3 Method

4.2.3.1 Participants

102 participants were recruited via Prolific with the same participation restrictions used in the previous experiments. Each experimental session took approximately 45-60 minutes, and participants were compensated $12 per hour. 8 participants were excluded due to less than 70% accuracy on comprehension questions. The analysis reported here includes data from 94 participants.

4.2.3.2 Materials

64 item sets were constructed. The design crossed three factors: Structure ((INT)egrated, ISO(lated)) x Grammaticality (GRAM, UNGRAM) x Match (MATCH, MISMATCH). Match referred to whether the experimental sentences involved coordinate structures containing matching ($N1_{PL}$-$N2_{PL}$) or mismatching ($N1_{PL}$-$N2_{SG}$) coordinates. The critical region, which included the auxiliary, either bore grammatical (plural) number marking, or ungrammatical (singular) marking. The coordinate subject was always embedded under *thinks*, *says*, *admits*, or *agrees*. A sample item set is provided in Table 4.9, where the critical region is bolded, and the spillover regions are underlined.

Experimental items were counterbalanced across 8 Latin-squared lists and pre-

Structure	Match	Grammaticality **critical region** | <u>spillover region</u>
INT	MATCH ($N1_{PL}$-$N2_{PL}$)	Ramona thinks that the chefs and the butchers quite quickly {GRAM: **have**, UNGRAM: ***has**} <u>prepared</u> an innovative menu for the restaurant.
	MISMATCH ($N1_{PL}$-$N2_{SG}$)	Ramona thinks that the chefs and the butcher quite quickly {GRAM: **have**, UNGRAM: ***has**} <u>prepared</u> an innovative menu for the restaurant.
ISO	MATCH ($N1_{PL}$-$N2_{PL}$)	Ramona thinks that the chefs, as well as the butchers, quite quickly {GRAM: **have**, UNGRAM: ***has**} <u>prepared</u> an innovative menu for the restaurant.
	MISMATCH ($N1_{PL}$-$N2_{SG}$)	Ramona thinks that the chefs, as well as the butcher, quite quickly {GRAM: **have**, UNGRAM: ***has**} <u>prepared</u> an innovative menu for the restaurant.

Table 4.9: Example item set from Experiment 5.

.

sented along with 44 fillers, for a total of 108 trials per experimental session. Half of the fillers were NP/Z garden path constructions, and the other half contained another type of focus-sensitive coordination construction that also featured obligatory prosodic breaks (*not only...but also*). All fillers contained coordination; half involved ordinary coordination, and the other half contained *as well as*-parentheticals within one of the two types of constructions mentioned above. Additionally, half of the fillers were grammatical, and the other half were ungrammatical (i.e., they contained agreement errors, like the UNGRAM conditions of the experimental items). Comprehension questions followed half of the experimental items and half of the fillers, and varied in which region of the sentence they targeted.

The experimental items featured a two word pre-verbal adverbial region in order to distance the edge of the second AWA-coordinate from the auxiliary. There were two reasons for this. The first is that previous reading studies on sentence-medial prosodic boundaries show inflated reading times on comma-marked words and facilitated times on immediately subsequent words (Hirotani et al., 2006). We aimed to prevent faster reading times on the critical region only because of this low-level reading tendency.

The second reason to add distance was because the processing of FSC in subject position has been understudied. If an attempt to engage in focus-sensitive ellipsis processing occurs at the earliest possible point, this might predict a penalty at the right edge of the second AWA-coordinate (56).

(56) Ramona thinks that the chefs $<t_1>$, as well as the butchers $<t_1>$, ...

One difference between these experimental items and previous investigations of FSC is that previous work primarily focuses on FSC in object position, in order to determine how the parser resolves ambiguity in the size of the remnant/correlate pair. Here, the parser cannot entirely resolve FS-ellipsis until processing of the remainder of the clause (*have prepared an innovative menu for the restaurant*) is complete. There are then two possibilities: (i) that the parser delays any operations related to resolving FS-ellipsis until the end of the sentence, at which point the elided content is reconstructed in both ellipsis sites (marked via $<t_1>$ in (56)), or (ii) that the parser begins an active search as in cataphoric dependencies (Frazier et al., 1983; Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007; Kroll, 2020) for the elided material. Typically, no reading time signature of such active search processes is present until the parser reaches a filled position or a potential antecedent that mismatches the properties of the left dependent. Under this view, we might not expect any early effect of FS-ellipsis. But, studies on active search have not previously investigated focus-sensitive constructions or constituents this large. If the resolution of FS-ellipsis is delayed, we would expect to see inflated sentence-final RTs for ISO conditions. If, on the other hand, there is some immediate "marking" of the position of the ellipsis site(s) immediately after processing of the second AWA-coordinate is complete, we might expect to see inflated RTs (relative to the INT conditions) on N2 (*butcher*), or on the next few words (the adverbial region). At present, we don't have enough information about the processing of these constructions to make more precise predictions about the exact process the parser engages in once processing of the second coordinate is complete. In any case, in order to ensure that the process of focus-sensitive ellipsis resolution did not interfere with reaction times on the critical region (the auxiliary), it was necessary to distance the edge of the coordinate two from this region. Ultimately, the current experiment does not evidence any reading time signature of resolving focus-sensitive ellipsis, so subsequent experiments will do away with the adverbial region included here.

If Keung and Staub (2018) are correct in concluding that linear proximity to the verb does not eliminate CCA effects, this additional distance should not pose a problem for the integrated conditions. One caveat, though, is that their distance manipulations only tested the likelihood of producing continuations consistent with the closer coordinate; they did not examine effects of distance during on-line comprehension. Thus, the current study aims to replicate their CCA finding in the processing of ordinary coordination, and to extend this to isolated AWA-coordination in order to further probe CSE and related hypotheses.

4.2.3.3 Procedure

The experiment utilized word-by-word, moving window bidirectional self-paced reading (B-SPR; Paape, Vasishth, Paape, and Vasishth, 2022), which allowed participants to progress or regress through the sentence by using the right or left arrow keys, respectively. In addition, participants could use the ESC key to return from the middle of the sentence immediately back to the beginning.

Participants were explicitly instructed to read with a natural intonation and to pause as needed for two reasons: first, to ensure that they adequately attended to positions where there were prosodic boundaries in the ISO conditions, and second, in order to counteract any disruption in the assignment of implicit prosodic structure potentially caused by reading one word at a time, which was a concern with the Maze task raised in §4.1. We expected to find that the comma-marked positions of the isolated conditions would lead to longer first-pass reading times than those same positions in ordinary coordinate structures, following Hirotani et al.'s (2006) finding that comma-marked regions induce wrap-up effects in early reading time measures[33].

One potential advantage of using B-SPR is that it offers the ability to collect reading time data more analogous to standard eye-tracking measures. For maximal comparability with Keung and Staub (2018), we analyzed first-pass reading times (the sum of all first-pass times on a region before any other region is visited), go-past times (the sum of all first-pass times on a position n and all preceding positions in the time period between the first visit on n and the first visit on anything to the right of n), and total reading times (the sum of all visit durations on a region) at the critical and spillover regions.

4.2.3.4 Predictions

The current experiment contrasts the predictions of Visibility, Context-Sensitive Retrieval (CSR), and Reinstantiation. Because I assume that AWA-coordinates are not backgrounded, I do not consider the role of Backgrounding here, but return to a discussion of this mechanism in the discussion (§4.2.5).

Firstly, we expect to replicate Keung and Staub's (2018) closest conjunct preference in INT structures: that RTs to GRAM-MATCH conditions should be faster than GRAM-MISMATCH conditions, and RTs to UNGRAM-MISMATCH conditions should be faster than UNGRAM-MATCH conditions. Following Keung and Staub, we expected to find evidence

[33] Our Maze data did not reflect this pattern; if anything, comma-marked ARC boundaries showed a speed-up relative to RRC boundaries in the Maze. One possible explanation for this is that comma-marked Maze foils may be easier to reject if they contain syntactic categories unlikely to be followed by a comma, for example.

of grammatical disruption in first-pass and go-past times and of ungrammatical facilitation in total times. Because each of the hypotheses make different predictions about how prosodically isolated content is expected to interact with the CCA effect, we frame the predictions in terms of relative degree of grammatical disruption and ungrammatical facilitation relative to the baseline INT structures. In (58)-(59) below, ✗ indicates predicted grammatical disruption, and ✔ indicates predicted ungrammatical facilitation. Inequalities signify predicted reaction times (RTs) relative to non-segmented structures.

Because Experiment 4, which investigated the processing of another focus-sensitive coordination structure, found evidence potentially consistent with Visibility, we consider its role in the current experiment as well. Under Visibility, the noun phrase in the current or most recent prosodic phrase should be most accessible. Because no potential retrieval candidates occupy the same prosodic phrase as the verb, accessing a previous phrase should be costly. Therefore, RTs to ISO conditions should be slower than INT conditions, contra the predictions of CSR. In addition, segmented structures should display a stronger grammatical disruption effect. Because $N2_{+SG}$ is the most accessible noun, and its retrieved context does not provide a good cue to $N1_{+PL}$'s encoding context, the mismatch between $N2_{+SG}$-V_{+PL} in ISO structures should lead to an even greater slowdown than the INT-MISMATCH condition. Therefore, while both ISO and INT should display evidence of grammatical disruption, ISO should show a greater RT penalty (57a). The ungrammatical facilitation effect should also be greater for ISO structures under Visibility: searching for $N2_{+SG}$ when it occupies a previous prosodic phrase should be more difficult than when it occupies the same phrase (57b). Therefore, both ISO/INT should display evidence of ungrammatical facilitation, but ungrammatical ISO-MISMATCH structures should display longer RTs than INT-MISMATCH ones, due to the cost of accessing a previous encoding context.

(57) Predicted RTs under Visibility

a. Grammatical disruption cost

✗ ✗

$\mathrm{N1}_{\mathrm{+PL}}$ $\mathrm{N2}_{\mathrm{+SG}}$ $\mathrm{V}_{\mathrm{+PL}}$ > $\mathrm{N1}_{\mathrm{+PL}}$ and $\mathrm{N2}_{\mathrm{+SG}}$ $\mathrm{V}_{\mathrm{+PL}}$

b. Ungrammatical facilitation cost

✔ ✔

* $\mathrm{N1}_{\mathrm{+PL}}$ $\mathrm{N2}_{\mathrm{+SG}}$ $\mathrm{V}_{\mathrm{+SG}}$ > * $\mathrm{N1}_{\mathrm{+PL}}$ and $\mathrm{N2}_{\mathrm{+SG}}$ $\mathrm{V}_{\mathrm{+SG}}$

Recall that CSR assumes that accessing a matching retrieval candidate in a segmented structure should be less costly than in a non-segmented one, because contextual interference from the contents of other segments should be reduced in the segmented case. Therefore, CSR predicts a diminished grammatical disruption effect for ISO conditions compared to INT. That is, $\mathrm{N2}_{\mathrm{+SG}}$ should not interfere to the same degree in ISO conditions, because the retrieval mechanism should be able to directly access the better matching candidate ($\mathrm{N1}_{\mathrm{+PL}}$), and the contents of the encoding context associated with N2 will not be co-activated. This is depicted in (58a). Conversely, CSR predicts a stronger ungrammatical facilitation effect. Because the features on the retrieval probe (+SG) can be used to directly access the medial encoding context containing N2, and because potential activation of $\mathrm{N1}_{\mathrm{+PL}}$ should be reduced in this case due to contextual dissimilarity, ISO structures should be more prone to illicitly retrieving N2. Therefore, RTs to ungrammatical ISO-MISMATCH conditions should be faster than to INT-MISMATCH. This is schematized in (58b). In addition, CSR may predict a main effect of structure such that RTs to ISO conditions are faster than to INT, because selecting a retrieval candidate in a segmented structure should give rise to less contextual interference than in a non-segmented structure.

(58) Predicted RTs under CSR

a. Grammatical disruption cost

$\boxed{N1_{+PL}}$ $\boxed{N2_{+SG}}$ $\boxed{V_{+PL}}$ $<$ $\boxed{N1_{+PL} \text{ and } N2_{+SG}\ V_{+PL}}$ ✗

b. Ungrammatical facilitation cost

* $\boxed{N1_{+PL}}$ $\boxed{N2_{+SG}}$ $\boxed{V_{+SG}}$ ✔ $<$ * $\boxed{N1_{+PL} \text{ and } N2_{+SG}\ V_{+SG}}$ ✔

Put another way, if retrieval can directly target a particular encoding context by using the context vector as a cue, this should allow the parser to reduce interference from an intervener in a non-targeted context. In grammatical conditions, this should allow the parser to effectively bypass the content contained within the medial encoding context, like we see for appositives. Therefore, a mismatching N2 should not disrupt agreement in grammatical, mismatching structures to the same degree as ordinary coordination. In ungrammatical conditions, this should allow the parser to more effectively access the illicit intervener, and thus facilitate agreement in ungrammatical, mismatching structures to an even greater degree than in ordinary coordination[34].

Reinstantiation assumes that in ISO structures, N1 should always become more accessible at the right boundary of the second coordinate, because its context is reinstated at this point. This predicts that within grammatical structures, RTs to ISO-MATCH/MISMATCH conditions should be equally fast, because $N1_{+PL}$ should always be more readily accessible prior to encountering the verb, which matches in features. Therefore, only the INT-MISMATCH condition should evidence grammatical disruption. This should result in faster RTs to ISO-MISMATCH than to INT-MISMATCH conditions, as depicted in (59a). Within grammatical conditions then, the predictions for Reinstantiation and CSR are identical. In the ungrammatical cases, although $N1_{+PL}$ will always become more accessible prior to processing the verb, the item-level +SG feature on

[34] One additional factor of note is that CSR does not limit ungrammatical facilitation to closest coordinate-matching configurations. The same speed up should be evident in *$N1_{+SG}$, as well as $N2_{+PL}$, V_{+SG}* structures. We do not test these structures here.

the verb should activate N2+SG, perhaps to the same degree in ISO and INT structures. Alternatively, if greater activation of $\text{N1}_{+\text{PL}}$ near the verb leads to greater difficulty accessing $\text{N2}_{+\text{SG}}$, this may result in reduced ungrammatical facilitation for ISO compared to INT. That is, ungrammatical ISO-MISMATCH should display slower RTs than INT-MISMATCH. This is schematized in (59b).

(59) Predicted RTs under Reinstantiation

a. Grammatical disruption cost

$\boxed{\text{N1}_{+\text{PL}}}$ $\text{N2}_{+\text{SG}}$ $\boxed{\text{V}_{+\text{PL}}}$ $<$ $\boxed{\text{N1}_{+\text{PL}} \text{ and } \text{N2}_{+\text{SG}}\ \text{V}_{+\text{PL}}}$

b. Ungrammatical facilitation cost

* $\boxed{\text{N1}_{+\text{PL}}}$ $\text{N2}_{+\text{SG}}$ $\boxed{\text{V}_{+\text{SG}}}$ $>=$ * $\boxed{\text{N1}_{+\text{PL}} \text{ and } \text{N2}_{+\text{SG}}\ \text{V}_{+\text{SG}}}$

We consider another possibility. If AWA-coordinates behave like appositives, we might expect matching parentheticals to yield less retrieval interference than matching ordinary coordination. In other words, this predicts a two-way interaction such that INT-MATCH conditions display longer RTs at the verb than INT-MISMATCH conditions, but ISO-MATCH/MISMATCH conditions are equally fast. Crucially, the grammatical disruption effect for integrated coordinates would have to fail to obtain in order to mirror the appositive relative clause bypassing effect; however, because grammatical disruption was observed by Keung and Staub (2018) in early eye-tracking measures and the current experiment uses self-paced reading, we consider the possibility that the measure in the current task (button press times) will not be sensitive enough to detect very early reading time effects. Furthermore, we expect that if a bypassing effect obtains, this should emerge in the grammatical conditions. This, too, would be consistent with the predictions of Reinstantiation. We have no specific prediction about whether or how bypassing should surface in ungrammatical conditions.

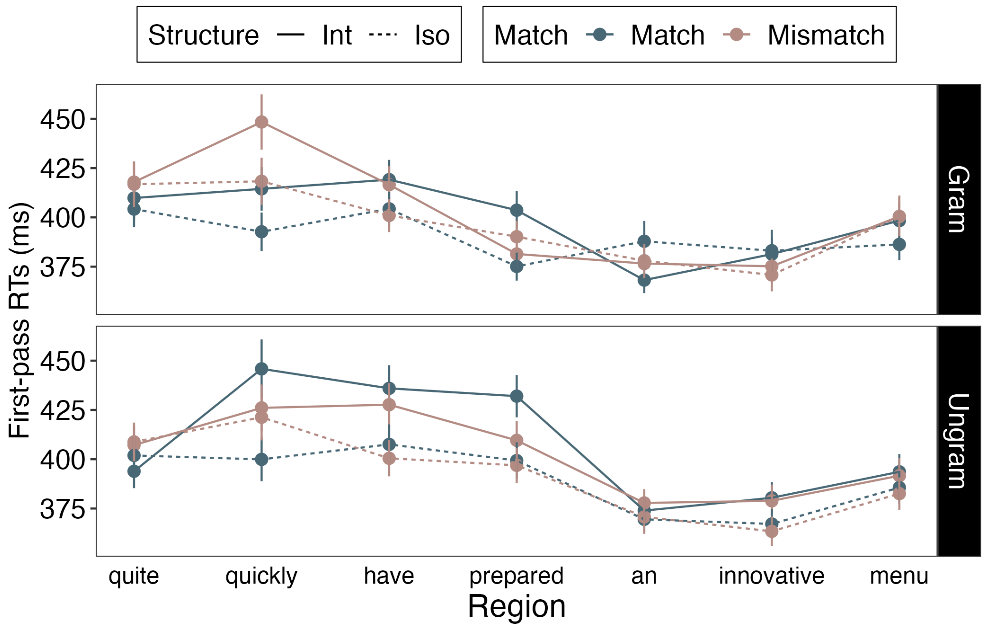
Overall, the main aims of the current experiment are to (i) determine whether other prosodically isolated structures, aside from appositives, display independence in their processing profile and (ii) delineate between the mechanisms discussed here, which each posit a different role for segmentation, and a different pattern for CCA in isolated structures.

4.2.4 Results

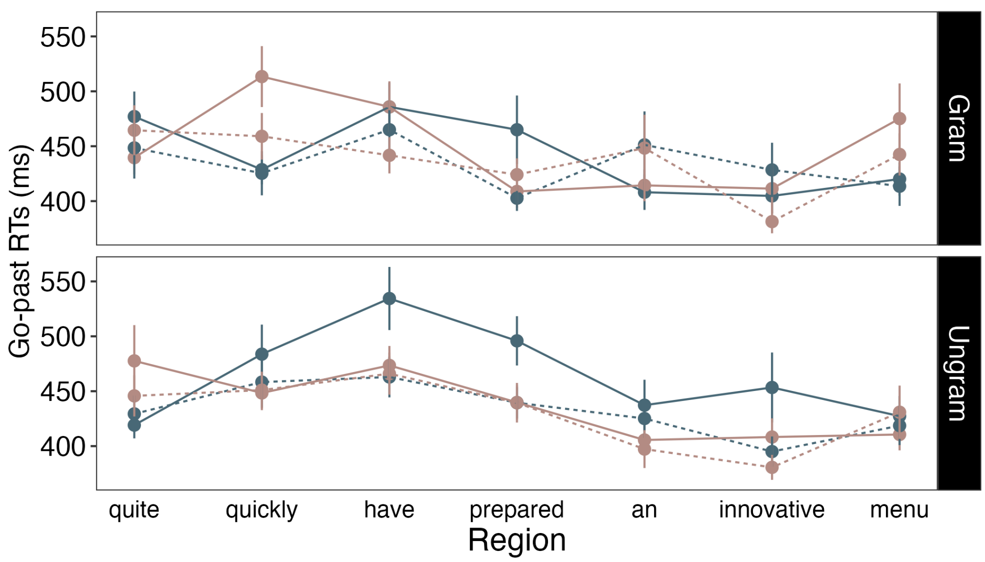
Average comprehension question accuracy was 82% (91% for the experimental items and 70% for the fillers). Lower comprehension accuracy for the fillers was likely due to the fact that some proportion of the questions following NP/Z fillers targeted the temporarily ambiguous region; this accords with previous investigations of NP/Z garden path constructions that have shown that comprehension question accuracy for the ambiguous region is generally lower due to the fact that the initial misinterpretation tends to linger.

Reading measures (first-pass, go-past, and total times) were computed using the `em2` package in R (Logacev & Vasishth, 2013). For the analysis of reaction times, Bayesian linear mixed-effects models[35] using the `brms` package (Bürkner, 2017) were fit to the critical and spillover regions using priors from Paape and Vasishth (2021). Because visual inspection of the full-sentence RTs suggested that differences between conditions began to emerge before the critical region, models were fit to the pre-critical (adverbial) region as well. Predictors were sum coded, with GRAM, ISO, and MATCH conditions coded as negative. $\widehat{R}$-values and posterior predictive checks indicated model convergence. Table 4.10 contains regression weights and 95% credible intervals from models fit to log RTs on each of the three regions. First-pass, go-past, and total times are plotted by region in Figure 4.3.

[35]`brm(logRT ~ Structure*Match*Grammaticality + (1 + Structure*Match*Grammaticality | Subject) + (1 + Structure*Match*Grammaticality | Item))`

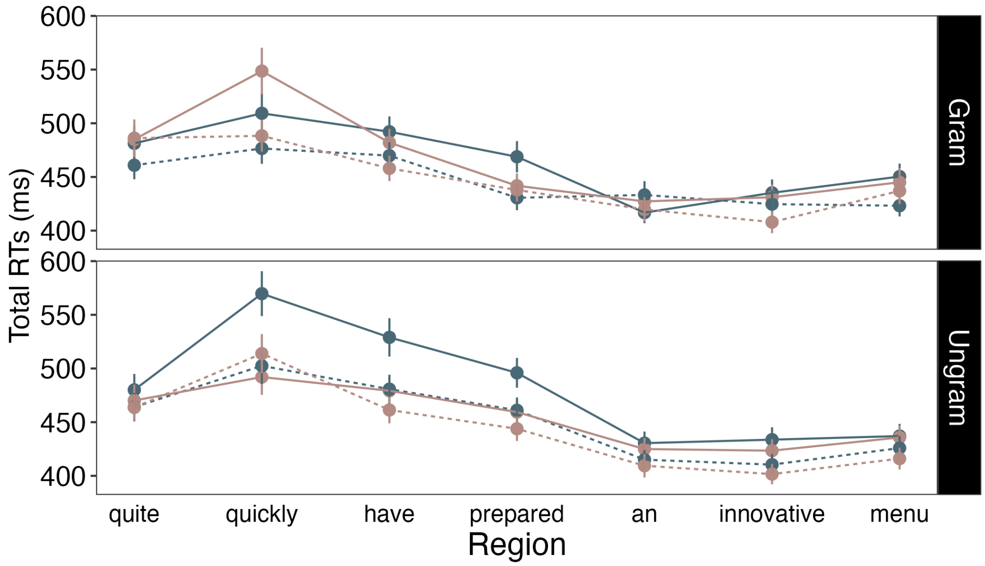


(a) First-pass times.



(b) Go-past times.

Figure 4.3: Word-by-word B-SPR latencies by condition for Experiment 5.



(c) Total times.

Figure 4.3: Word-by-word B-SPR latencies by condition for Experiment 5. (cont.)

I first summarize the model results for first-pass SPR times. At the critical ($\hat{\beta}$ = 0.04, CrI = [0.02, 0.04]) and spillover ($\hat{\beta}$ = 0.02, CrI = [0.00, 0.04]) regions, there was a main effect of Structure such that INT conditions were read slower. This is consistent with the observation that post-comma regions are typically read faster (Hirotani et al., 2006). No other credible effects emerged at the critical region. In the spillover region, RTs to UNGRAM conditions were credibly slower than those to GRAM conditions ($\hat{\beta}$ = 0.03, CrI = [0.01, 0.06]). In addition, there was an interaction of Structure and Match ($\hat{\beta}$ = -0.05, CrI = [-0.09, -0.01]), such that MATCH conditions were read slower in INT structures only. This affirms the prediction that isolated content interferes during retrieval to a lesser degree than the same content in integrated coordinate structures, and is analogous to the ARC-bypassing effects reported elsewhere (Dillon et al., 2017; S. Kim & Xiang, 2023). At the pre-critical region, there was a main effect of Structure ($\hat{\beta}$ = 0.04, CrI = [0.02, 0.07]), such that RTs to INT conditions were read slower than RTs to ISO conditions, and a main effect of Match ($\hat{\beta}$ = 0.03, CrI = [0.00, 0.05]) , such

First-Pass RTs	Pre-critical (*quickly*) $\hat{\beta}$	CrI	Critical (*have*) $\hat{\beta}$	CrI	Spillover (*prepared*) $\hat{\beta}$	CrI
Struc	**0.04**	**(0.02,0.07)**	**0.04**	**(0.02,0.06)**	**0.02**	**(0,0.04)**
Match	**0.03**	**(0.00,0.05)**	-0.01	(-0.03,0.01)	-0.01	(-0.02,0.01)
Gram	0.01	(-0.01,0.03)	0.01	(-0.01,0.03)	**0.03**	**(0.01,0.06)**
Struc*Match	-0.02	(-0.07,0.02)	0.00	(-0.04,0.04)	**-0.05**	**(-0.09,-0.01)**
Struc*Gram	0.00	(-0.04,0.05)	0.03	(-0.01,0.07)	0.03	(-0.01,0.06)
Match*Gram	-0.02	(-0.06,0.02)	-0.00	(-0.05,0.04)	-0.02	(-0.06,0.02)
Struc*Match*Gram	-0.04	(-0.12,0.05)	-0.00	(-0.08,0.08)	0.03	(-0.05,0.11)
Go-Past RTs	$\hat{\beta}$	CrI	$\hat{\beta}$	CrI	$\hat{\beta}$	CrI
Structure	**0.04**	**(0.01,0.06)**	**0.04**	**(0.01,0.07)**	**0.02**	**(0.00,0.05)**
Match	**0.03**	**(0.00,0.06)**	-0.02	(-0.05,0.00)	-0.01	(-0.04,0.01)
Gram	0.01	(-0.01,0.04)	0.01	(-0.02,0.04)	**0.04**	**(0.02,0.07)**
Struc*Match	-0.01	(-0.06,0.05)	-0.00	(-0.06,0.05)	**-0.06**	**(-0.11,-0.01)**
Struc*Gram	-0.01	(-0.06,0.04)	0.02	(-0.03,0.08)	0.03	(-0.02,0.08)
Match*Gram	-0.04	(-0.09,0.01)	-0.00	(-0.06,0.05)	-0.03	(-0.08,0.03)
Struc*Match*Gram	-0.02	(-0.14,0.09)	-0.05	(-0.16,0.07)	0.02	(-0.07,0.12)
Total RTs	$\hat{\beta}$	CrI	$\hat{\beta}$	CrI	$\hat{\beta}$	CrI
Structure	**0.04**	**(0.01,0.08)**	**0.04**	**(0.02,0.06)**	**0.03**	**(0.01,0.06)**
Match	0.00	(-0.03,0.02)	**-0.03**	**(-0.05,-0.01)**	-0.02	(-0.04,0.01)
Gram	0.02	(-0.01,0.05)	0.01	(-0.01,0.03)	**0.04**	**(0.01,0.07)**
Struc*Match	-0.03	(-0.09,0.02)	-0.00	(-0.05,0.04)	**-0.05**	**(-0.09,-0.00)**
Struc*Gram	-0.02	(-0.07,0.04)	0.01	(-0.03,0.06)	0.02	(-0.03,0.06)
Match*Gram	**-0.06**	**(-0.11,-0.00)**	-0.02	(-0.07,0.02)	-0.04	(-0.09,0.02)
Struc*Match*Gram	-0.09	(-0.21,0.02)	-0.02	(-0.11,0.07)	0.03	(-0.06,0.12)

Table 4.10: Bayesian linear mixed-effects models fit to log RTs at the pre-critical, critical, and spillover regions of Experiment 5.

that RTs for MISMATCH conditions were slower than those for MATCH conditions.

In go-past times, a main effect of Structure emerged at the critical ($\hat{\beta}$ = 0.04, CrI = [0.01, 0.07]) and spillover ($\hat{\beta}$ = 0.02, CrI = [0.00, 0.05]) regions. Additionally, there was a main effect of Grammaticality at the spillover region (UNGRAM > GRAM; $\hat{\beta}$ = 0.04, CrI = [0.02, 0.07]), and an interaction of Structure and Match ($\hat{\beta}$ = -0.06, CrI = [-0.11, -0.01]) such that INT-MATCH conditions were read slower than ISO-MATCH conditions. Main effects of Structure (INT > ISO; $\hat{\beta}$ = 0.04, CrI = [0.01, 0.06]) and Match (MISMATCH

> MATCH; $\hat{\beta}$ = 0.03, CrI = [0.00, 0.06]) emerged at the pre-critical region.

In total times at the critical region, there was a main effect of Structure (INT > ISO; $\hat{\beta}$ = 0.04, CrI = [0.02, 0.06]) and a main effect of Match (MISMATCH > MATCH; $\hat{\beta}$ = -0.03, CrI = [-0.05, -0.01]). In the spillover regions, there were main effects of Structure (INT > ISO; $\hat{\beta}$ = 0.03, CrI = [0.01, 0.06]) and Grammaticality (UNGRAM > GRAM; $\hat{\beta}$ = 0.04, CrI = [0.01, 0.07]). In addition, there was an interaction between Structure and Match ($\hat{\beta}$ = -0.05, CrI = [-0.09, -0.00]), once again, such that INT-MATCH conditions were slower than ISO-MATCH conditions. The pre-critical region showed a main effect of Structure (INT > ISO; $\hat{\beta}$ = 0.04, CrI = [0.01, 0.08]) and an interaction between Match and Grammaticality; $\hat{\beta}$ = -0.06, CrI = [-0.11, -0.00], such that UNGRAM-MATCH conditions were slower than GRAM-MATCH conditions and GRAM-MISMATCH conditions were slower than UNGRAM-MISMATCH conditions.

Because differences began emerging at the pre-critical position, we investigated this further in a post-hoc analysis. Across all three measures, there was a trend towards the predicted three-way interaction in the pre-critical region (note that the 95% CrIs for the three-way interaction indicate that the probability mass of the posterior distribution is not centered around 0). This emerged most clearly in total times. Models fit to the pre-critical region revealed that this trend was not credible in any measure. A Bayes Factor analysis comparing 3-way interaction vs. no interaction models (Wagenmakers et al., 2018) on total times suggested moderate evidence for the non-interaction model (BF_{10} = 0.16).

Finally, we fit `brms` models[36] to the ISO and INT conditions independently for total times, where the trending 3-way interaction was most evident, in order to determine whether (i) the two-way interaction predicted for standard CCA in INT structures obtained and (ii) this interaction disappeared in ISO structures. The results of this

[36] `brm(logRT ~ Match*Grammaticality + (1 + Match*Grammaticality | Subject) + (1 + Match*Grammaticality | Item))`

post-hoc analysis are in Table 4.11.

	INT Model		ISO Model	
Total RTs	$\hat{\beta}$	CrI	$\hat{\beta}$	CrI
Match	-0.02	(-0.06,0.02)	0.01	(-0.03,0.05)
Gram	0.01	(-0.03,0.05)	0.03	(-0.01,0.07)
Match*Gram	**-0.1**	**(-0.18,-0.02)**	-0.01	(-0.08,0.06)

Table 4.11: By-structure (INT-only and ISO-only) Bayesian linear mixed-effects models fit to total log RTs at the pre-critical region of Experiment 5.

The INT-only model revealed a credible interaction between Match and Grammaticality ($\hat{\beta}$ = -0.01, CrI = [-0.18,-0.02]), validating the basic CCA pattern. The ISO-only model revealed no credible effects. This contrast provides suggestive evidence in favor of a difference between ISO vs. INT, such that CCA effects are diminished for ISO structures. Therefore, I discuss this trend further in the discussion.

4.2.5 Discussion

The results replicate Keung and Staub's (2018) CCA pattern for INT structures in total times at the pre-critical region only. In our stimuli, the two-word adverbial region preceding the auxiliary (critical) and main verb (spillover) regions may have encouraged participants to initiate retrieval of the subject early.

Overall, the facilitation for ISO structures at the pre-critical and spillover regions is most consistent with Reinstantiation. However, the results do not align exactly with the predictions detailed in §4.2.3.4, because differences between ISO vs. INT structures began emerging prior to the critical region. Recall that the results provided suggestive support for a the 3-way interaction at the pre-critical region in total times. While the grammatical disruption and ungrammatical facilitation effects emerged in INT structures, both effects were eliminated in ISO structures. Across-the-board facilitation for ISO structures seemingly aligns with the predictions outlined for CSR, but because these differences began emerging prior to the auxiliary, which bore number features that CSR must rely on in order to initiate retrieval, the effects at the adverbial are better explained by Reinstantiation. If prosodic segmentation generally leads to reduced contextual interference, and the right edge of the prosodic phrase containing the AWA coordinate leads to reinstatement of the initial encoding context, $N1_{+PL}$ should be "pre-activated" at the pre-critical region, even prior to encountering features on the retrieval probe that guide the search for a subject. This suggests that the processing load of subject retrieval is alleviated for ISO structures, because the work is partially done in advance. I return to this argument shortly.

The results at the spillover region further support Reinstantiation. This is most evident in GRAM conditions. There, we saw evidence of a bypassing effect for AWA-coordinates, where the effect of retrieval interference surfaced only for INT conditions (GRAM-MATCH > GRAM-MISMATCH), not ISO ones. This pattern reliably emerged across all three reading measures. This suggests that prosodically segmented structures benefit from reduced interference, even when the content of an isolated phrase is not discourse-subordinate. In UNGRAM conditions, the fact that ISO structures show facilitation could be interpreted as evidence for CSR. In the ISO-MISMATCH case, accessing $N2_{+SG}$ may facilitate processing due to the availability of a feature-matching retrieval candidate that is contained within a distinct encoding context, and should thus engender less contextual interference. It's somewhat mysterious that an ungrammaticality penalty does not emerge for ungrammatical ISO-MATCH structures, though. In this condition, there is no item that matches the number features on the auxiliary. Therefore, we expected to observe some cost, though perhaps a smaller one than in INT structures. None of the proposed mechanisms can account for this. It's possible that readers were not processing the ISO structures as deeply, though we think this unlikely because comprehension question accuracy was quite high for the experimental items. At present, we do not have an explanation for the lack of a difference between ungrammatical ISO-MATCH/MISMATCH.

Nevertheless, the results (i) overwhelmingly support Reinstantiation because AWA-coordinates appear to be bypassed and (ii) are inconsistent with the predictions of Visibility because we do not find evidence of difficulty associated with ISO structures, or strengthened CCA effects for ISO structures.

Recall that bypassing of ARCs has been previously observed for filler-gap dependencies and pronominal anaphora in sentences containing appositive relative clauses (ARCs). For example, S. Kim and Xiang (2023) find that anaphoric resolution in (60a) is less costly than (60b). Recall that they argue that the subordinate discourse status of ARCs drives their bypassing, per the Backgrounding hypothesis introduced in §3.1.3.

(60) a. **The violinists**, who admired the **the singers**, invited **their mentors**...

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

If ARCs are reliably interpreted as subordinate units, the bypassing effect in (60a) may have obtained simply because discourse-based dependencies like anaphora should be sensitive to the discourse structure of a sentence, which is argued to influence accessibility of antecedents for pronoun resolution (Asher & Lascarides, 2003; Polanyi, 1988). However, the current experiment extends this finding to agreement in non-subordinate, sentence-medial prosodic phrases. Because ARCs and AWA-coordination pattern together in this respect, the bypassing effect cannot be attributed to the discourse status of appositives.

Another difference between S. Kim and Xiang's study and ours is that they used ambiguous constructions. The anaphor in (60) could refer to either *the violinists* or *the singers*, and so the parser may choose to retrieve either antecedent and proceed with parsing the rest of the sentence without issue. In our case, some representation containing both nouns of a complex subject must eventually be retrieved from memory in order to fully resolve coordinate agreement and integrate the subject with the verb. In integrated coordinate structures, where both coordinands are encoded as contextually similar due to structural and temporal proximity, if the retrieval mechanism probabilistically selects one of the nouns at an early stage, this should not necessarily pose a problem later on, because similarity in the contextual state of the retrieved item should allow the other coordinand to become co-active, thus facilitating coordinate resolution. In isolated coordinate structures, however, it remains unclear whether both coordinates are ever accessed completely, because they are encoded in distinct contexts and thus the retrieved context of one does not provide a valid cue to the contextual state of the other. By the same process, initial access to one coordinand may inhibit access to a prosodically separated second coordinand if it cannot become co-active as easily, due to contextual distinctiveness. In this case, some global structural representation must be activated in order to perform later processes on the complete subject. Such a mechanism would predict downstream consequences; this is the focus of Chapter 5.

There is another important difference between sentence-medial ARCs vs. AWA-coordination in their discourse representation. AWA-coordinates comprise independent discourse units in their sentence-final representation, but when they are in subject-position as they were in the current experiment, they are not complete discourse units at the time of processing the verb, the region where we saw evidence of bypassing. At most, the parser may mark the position of the ellipsis site following each of the foci in these structures, for the purpose of reconstructing focus-sensitive ellipsis later on, as in (61).

(61) (The chefs <t_1>) (as well as the butchers <t_1>) (have prepared…)

Against this backdrop, bypassing cannot be attributed to a process that depends on the completion of a particular discourse unit, as has been suggested for ARCs (c.f.

Dillon et al., 2017; Duff et al., 2023).

Because Reinstantiation offers a more parsimonious account of bypassing, I conclude that it is better equipped to explain the processing of both structures than assuming ARCs are always bypassed because they are backgrounded. On that basis, I rule out the Backgrounding hypothesis introduced in §3.1.3. The results of the current experiment suggest that incremental reinstantiation of previous encoding contexts may be a more general property of sentence-medial intonational phrases.

Mechanistically, Reinstantiation suggests that upon completion of a sentence-medial prosodic phrase, the right boundary triggers reactivation of the contextual state associated with the initial encoding context. This process may be akin to the "attentional refreshing" of memory traces proposed in Time-Based Resource Sharing models (Barrouillet & Camos, 2021), where additional time between processing and encoding events in complex reading span tasks has been shown to strengthen memory representations, not degrade them. If readers use prosodic boundary positions as opportunities to dwell in order to refresh previous encoding contexts, this may serve to strengthen their representation in long-term memory, provide a means of constructing a global linguistic structure, or integrate linguistic content across distinct encoding contexts for the purpose of establishing discourse and conceptual relationships. In support of this idea, reading times for ISO structures were longer than INT structures at the right boundary of the AWA-coordinate across all reading measures. First-pass times are plotted in Figure 4.4, where ISO > INT at the right boundary (*butchers,*) and ISO = INT at the left boundary (*chefs,*), except in the GRAM-MISMATCH panel where the left ISO boundary was also longer. The same asymmetry in left vs. right boundary dwell times has been reported for ARCs in eye-tracking (Hirotani et al., 2006). Experiment 6 explores these ideas further by investigating the types of linguistic processes at work at sentence-medial boundary positions.

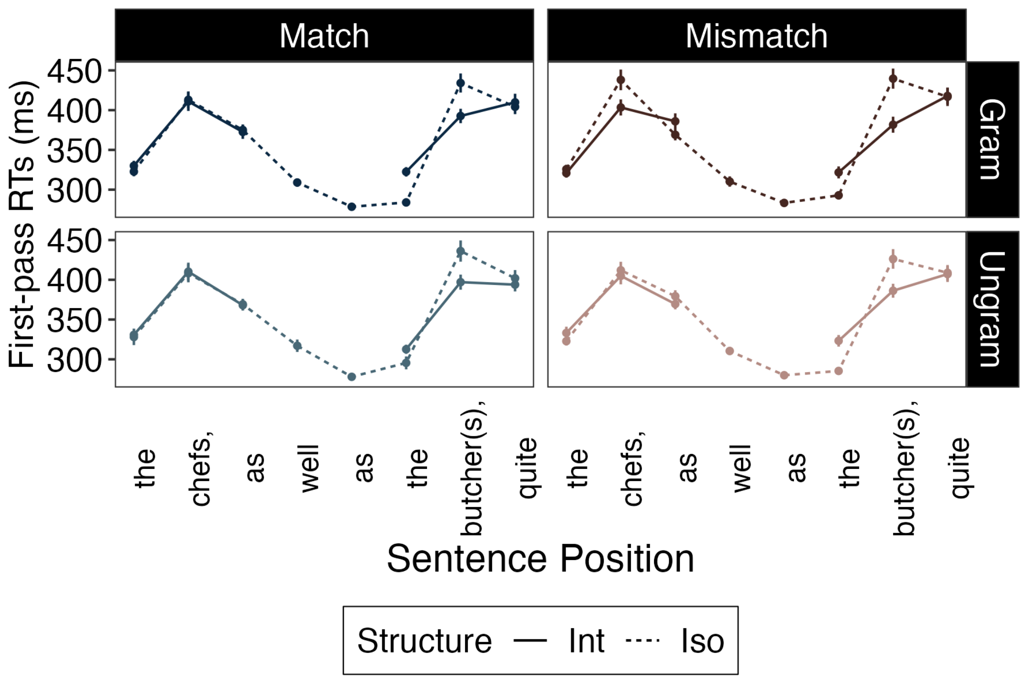


Figure 4.4: First-pass latencies by condition for boundary positions in Experiment 5.

As proposed for ARCs in §3.1.3, I suggest that Reinstantiation yields the timecourse for isolated AWA processing in (62).

(62) Reinstantiation for AWA-coordination

1 2 3

The chefs, | as well as the butchers, | have prepared an innovative menu

1 : *chefs,* → context shift

2 : *butchers,* → context shift, reinstate initial context

3 : *prepared* → retrieve subject, with initial context more active

If encapsulated prosodic phrases encourage periodic reinstantiation of earlier content, this would explain why constructions with a single medial prosodic boundary, like *not only...but also* (NOBA), diverge in their processing profile from ARCs and AWA-coordination. Reinstantiation is a process that is hypothesized to occur once the initial parsing of a prosodically bounded sentence-medial unit is complete – that is, once the left and right prosodic boundaries have been encountered. At this point, it may also make the prediction that there is some "loading cost" to reinstating a previous context (c.f. Oberauer, 2005, 2009), at the right edge of the current unit (the dwell time pattern in the current experiment and the results of Experiment 6 align with this idea). In the case of a single sentence-medial boundary, like in NOBA, Reinstantiation is not hypothesized to be operative at the right edge of the initial coordinate (which also marks the first segment of the sentence). Upon entering the second coordinate and triggering the retrieval of a dependency (like in Experiment 4), if there is a cost associated with accessing a distinct encoding context, this will surface as a "visibility" effect, the source of which is contextual reinstatement. This would explain the penalty for *one*-anaphora in NOBA structures that we observed in Experiment 4, if this was indeed due to Visibility (see §4.1 for discussion). If true, this should predict bypassing for NOBA structures in subject position, like in (63).

(63) Ramona heard that not only the chefs, but also the butchers, were preparing an innovative menu.

In order to rule out an alternative interpretation of the results, I conclude this section by discussing another notable property of the experimental design. The current study compared the processing of ordinary coordination with focus-sensitive coordination. Previous work has established that the processing of focus is unique. Cues to upcoming semantic foci have been shown to guide attention allocation (Cutler & Fodor, 1979), foci are remembered better than non-foci (S. L. Birch & Garnsey, 1995), and participants perform better on change (Sturt, Sanford, Stewart, & Dawydiak, 2004) and error (Bredart & Modolo, 1988) detection decisions on focused positions. Foci are also more accurately retrieved from memory in speed-accuracy tradeoff studies (Foraker & McElree, 2007). It is therefore argued that foci are privileged in memory, because they are attended to and encoded more deeply. It then also stands to reason that foci may be more accessible during retrieval operations, because they benefit from higher activation at the time of encoding.

In the ISO conditions of the present experiment, focus-sensitivity of the structure should already be established by the time the critical region is encountered, and the retrieval candidates themselves are contrastive foci. This is not the case in the INT conditions. Therefore, the general advantage for ISO structures may be driven not by the fact that post-comma positions are read faster, but instead because retrieving foci from memory may be easier than retrieving non-foci. The fact that ISO-MATCH conditions exhibit less retrieval interference could then be explained not by a tendency to bypass distinct contexts, but instead by generally facilitated access to either of the coordinates when they are focused. The results of Experiment 4 can partially inform the validity of this explanation. There, anaphoric resolution times were slower, not faster, in *not only...but also* constructions (64a), where the antecedent was focused, compared to *...and also...* constructions (64b), where the fact that the antecedent is focused cannot be determined until encountering *also* in the second coordinate, at the earliest. However, if the focus structure of the first coordinate is retroactively established following the processing of the coordinator (*and also*), it's possible that the focus structure of the first coordinate is established for both structures by the time the parser reaches the anaphor.

(64) a. Imala trusted *not only* **the inventor** with fifty gizmos, *but also* **the one**...

b. Imala trusted the inventor with fifty gizmos, *and also* the one...

One way to interpret the contrast between the results of Experiments 4 and 5, then, is that when focus structure is held constant, the effect of Visibility emerges (Exp. 4), but given the contrast between a focus-sensitive vs. non-focus-sensitive construction (Exp. 5), the effect of Visibility is rendered inert. In other words, perhaps the effect of focus supersedes the effect of boundaries. There are several reasons to think this is not the case. First, this would be inconsistent with studies that find that the presence of prosodic boundaries mediates the effect of focus in attachment decisions (Van Handel, 2022). Second, the fact that facilitation of the type we observe in the current experiment also obtains for appositive relative clauses, which are not focus-sensitive constructions (although they certainly introduce contrasting material in our stimuli), suggests that there is something else at work here.

A follow-up experiment could replicate the current study with a minimally different focus-sensitive construction like *and also*, which as discussed in §4.2.1, is information structure- and meaning-identical with *as well as* (Hulsey, 2008). This could be achieved via the contrast in (65).

(65) a. Ramona thinks that the chefs, as well as the butchers, have prepared an innovative menu for the restaurant.

b. Ramona thinks that the chefs and also the butchers have prepared an innovative menu for the restaurant.

Experiment 6 takes a different approach. It contrasts prosodically isolated vs. integrated AWA-coordination, two identical focus-sensitive structures, rather than using ordinary coordinate structures as a baseline. The results of this experiment rule out a focus-based explanation of Experiment 5, and also serve to validate the particular role of prosodic boundaries in triggering context shifts.

4.3 Experiment 6

The proposed timecourse of processing AWA-coordination put forward in the previous section makes a particular claim about the processes active at sentence-medial prosodic boundaries. The purpose of Experiment 6 is to validate the proposed timecourse, and to corroborate the claim that the mechanism responsible for shifting encoding contexts is sensitive to the presence of prosodic boundaries in particular. A secondary aim is to explore the relationship between dwell times and the corresponding memory representation of linguistic segments. We sought to determine whether longer processing times at prosodic boundary positions of isolated AWA-coordination (compared to its integrated counterpart) would lead to a corresponding increase in sensitivity (in Signal Detection Theory terms) to the content of each coordinate. In order to investigate this, the following section reviews the literature on clause wrap-up effects – the tendency for readers to dwell longer on clause boundaries marked with punctuation – then considers how wrap-up processes may interact with the context-sensitive memory mechanisms proposed in the current dissertation. Experiment 6 contrasts the predictions of Context-Sensitive Encoding and Reinstantiation, but also returns to a discussion of Compression. The results reported here ultimately provide additional evidence in favor of Context-Sensitive Encoding and Reinstantiation, and in particular, the role of prosodic boundaries in triggering encoding context shifts.

4.3.1 Wrapping Up at Boundaries

Readers regularly dwell longer at clause boundaries that are marked with punctuation than those that are unmarked in self-paced reading (Just, Carpenter, & Woolley, 1982) and eye-tracking (Rayner, Kambe, & Duffy, 2000). This has been termed the clause *wrap-up effect* (Just & Carpenter, 1980). The existence of wrap-up effects has been a long-standing mystery in psycholinguistics and the psychology of reading. Decades of work on sentence processing has established that syntactic and semantic processing occurs on an incremental, word-by-word basis, and yet, readers tend to dwell on certain boundary positions. Many investigations of wrap-up effects have noted that it is unclear what linguistic work remains to be done at marked clause boundaries. The previous sections of this dissertation have suggested several possibilities related to both memory and linguistic processes: (i) that boundaries lead to shifts in the temporal context vector (though it is unclear whether this process consumes a significant amount of processing resources), (ii) that previous encoding contexts can be reinstated at boundaries, (iii) and that discourse processes may apply at boundary positions, including updating of the discourse status of a particular unit, or establishing a discourse relationship between units. At least suggestion (iii) is consistent with previous work on clause-final wrap-up effects, suggesting that whatever clause wrap-up is, it is linguistically relevant.

But ultimately, the source and function of clause wrap-up remains somewhat mysterious. Some researchers have suggested that wrap-up effects reflect an oculomotor response to the low-level visual features of punctuation (Hill & Murray, 2000; Warren, White, & Reichle, 2009). However, the fact that wrap-up effects principally obtain in go-past reading times in eye-tracking, not first-pass times, challenges this reasoning (S. Andrews & Veldre, 2021). Go-past times index re-reading and are typically taken to reflect more involved linguistic processing. Furthermore, different types of punctuation yield different wrap-up costs: comma-marked words yield smaller wrap-up effects than sentence-boundaries marked with a period. If wrap-up times were modulated by visual complexity during reading alone, we might expect the opposite pattern of results, because commas are visually more complex than periods. Previous work has also established that wrap-up cost is not modulated by the syntactic complexity of a clause (Warren et al., 2009), suggesting that it cannot stem from delaying certain syntactic/semantic processes due to processing bottlenecks during incremental comprehension, for example.

Other studies have found promising evidence suggesting that clause-final wrap-up is tied to linguistic and metacognitive processes. These studies variably attribute wrap-up effects to (i) long-term memory processes, (ii) the assignment of implicit prosodic structure, (iii) conceptual integration across segments, or (iv) individual differences in reading strategy. Evidence for each of these sources is summarized briefly below.

❧ Wrap-up as working memory → long-term memory "consolidation"

Several studies suggest that wrap-up costs reflect the relocation of linguistic content from the working memory buffer to long-term memory, in order to free up processing resources for the incoming segment. There are two ways to think about the consequences of this relocation. On the one hand, one could posit that this process leads to decay of previous material (Hirotani et al., 2006). This claim is based on the observation that readers are less likely to make regressive eye movements passed a preceding prosodic boundary, perhaps because the contents of earlier phrases are less accessible in memory. As discussed in §2.2.1, this conception is at odds with the beneficial effect of segmentation on sentence-final representations.

Alternatively, dwelling at boundaries has been argued to strengthen the representation of a text in memory. This brings to mind several findings from the domain-general memory literature. The first finding relates to the Time-Based Resource Sharing (TBRS) model (Barrouillet & Camos, 2021), which suggests that additional processing time leads to the strengthening of decaying memory traces through an attentional focusing and refreshing mechanism. In other words, when participants are afforded more time between processing events, they use this as an opportunity to re-attend to previously encoded content so that it is not lost. This process is reminiscent of the Reinstantiation mechanism we've proposed here[37]. The second finding relates to the notion of *consolidation* in memory, which has been shown to occur at very long timescales (e.g., during sleep). Consolidation is the process by which "a temporary, labile memory is transformed into a more stable, long-lasting form" (Squire, Genzel,

[37] But note that the definition of Reinstantiation assumes that contextual reinstatement is (in part) linguistically driven, e.g., by the need to establish a subject-verb relationship across prosodically disjointed content.

Wixted, & Morris, 2015). The issue with applying this logic to sentence processing is that the majority of our effects of interest unfold on the order of a few hundred milliseconds. However, some domain-general work points to a role for "short-term memory consolidation", on the basis of studies that find longer retention intervals to be beneficial to memory representations. Finally, Sols, DuBrow, Davachi, and Fuentemilla (2017) find EEG evidence for the rapid (~200-800 ms) reinstatement of previous events at encoding context boundaries, although they claim that this is representative of reinstatement of the just-encoded context, not a previous one. Together, these findings suggest that effortful processing at boundaries may reflect the process of forming a more durable memory representation; this has also been previously proposed as a general function of prosodic boundaries (Slowiaczek & Clifton, 1980).

To our knowledge, only two wrap-up studies have explicitly tested this idea. Haberlandt, Graesser, Schneider, and Kiely (1986) found that sentence-final dwell times positively correlate with recall accuracy for sentences. They attribute this to transfer of linguistic content into "more permanent text memory." L. Miller and Stine-Morrow (1998) similarly found that longer wrap-up times in sentence-medial and final positions were associated with higher recall accuracy, particularly for groups of participants that were provided with less topical information about the text beforehand. Crucially, these results refute the claim by Hirotani et al. (2006) that wrap-up processes induce decay of previous material.

❧ Wrap-up as an implicit prosodic reflex

Hirotani et al. (2006) explore the idea that wrap-up effects are a reflex of assigning an implicit prosodic structure (J. Fodor, 1998, 2002a, 2002b) to sentences, because intonational phrase boundaries involve pauses and phrase-final lengthening in speech. They find that readers reliably dwell on the boundaries of vocative and parenthetical constructions, as in (66)-(67), but only dwell on the right boundary of appositives (68).

(66) John, go to the library for me.

(67) Fred searched all the closets, Anabelle indicated.

(68) Those men, who live in New Haven, go to Yale University.

Recall from §3.1.1.3 that in speech, appositives consistently feature pauses and boundary tones at both their left and right boundaries. Therefore, the fact that readers do not tend to dwell on left ARC boundaries cannot be a consequence of implicit prosodic assignment alone, if the implicit prosody for ARCs tracks their overt prosody. In a second experiment, the authors found no reliable effects of commas for ARCs. They suggest that intonational phrase boundaries influence the presence of wrap-up, but not-at-issue material receives less attention during reading (c.f. Dillon et al., 2014). If appositives are attended to less, this should result in worse encoding of their contents in memory. Note that this explanation is inconsistent with the results of Experiment 1, which found marginally better memory for ARC content. However, work on individual differences presents an alternate explanation for the variability in dwell time effects at boundaries (see the discussion of Stine-Morrow et al., 2010, below).

❧ Wrap-up as conceptual updating

Early work suggested that wrap-up effects index the cost of integrating segments into a discourse representation or situation model. Specifically, Just and Carpenter (1980) propose that wrap-up relates to establishing inter-clause relations, and also other discourse processes (like establishing co-reference). As mentioned earlier, semantic and discourse processes, like referential processing, are generally not delayed until the clause-final position (see Stowe, Kaan, Sabourin, and Taylor, 2018, for other examples). However, it's very plausible that segment-level discourse processes, like forming connections between segments in a situation model (Zwaan & Radvansky, 1998) or establishing coherence relations in the discourse representation (Asher & Lascarides,

2003) are processes that must wait until the processing of clausal/propositional information is complete. In support of this proposal, Millis and Just (1994) found that subordinate *because*-clauses in sentence-final position yield stronger sentence-final wrap-up effects compared to independent clauses not linked via subordinators. Relating this finding to work suggesting that *because*-clauses are better remembered than *and/or*-clauses (Caron, Micko, & Thüring, 1988), the authors put forward the Reactivation Hypothesis: that at the end of the second clause, the first clause is reactivated in memory and integrated with the second clause for the purpose of constructing a relationship between the clauses. Furthermore, strengthening the conceptual relatedness between clauses also increased the size of the sentence-final wrap-up effect. This account suggests that wrap-up may be a multi-pronged process that employs memory- and discourse-relevant operations. I return to this suggestion shortly.

❧ Wrap-up as self-regulated resource allocation

Several studies link wrap-up effects to the task demands of experimental contexts, as more involved tasks typically yield larger wrap-up costs. For example, Haberlandt et al. (1986) found that sentence-final wrap-up costs are increased for sentence recall tasks compared comprehension and "free reading" tasks. Weiss, Kretzschmar, Schlesewsky, Bornkessel-Schlesewsky, and Staub (2018) found that comprehension question difficulty influences the size of sentence-final wrap-up effects, and that larger wrap-up costs are driven by re-reading behavior, not first-pass reading times. Their difficult comprehension questions targeted readers' ability to successfully engage in syntactic reanalysis (see §5.2 for more detail). Although wrap-up times were modulated under these conditions, longer times did not lead to better comprehension accuracy. As such, they suggested that wrap-up costs reflect a checking mechanism that responds to the demands of the task, not reanalysis processes themselves. Similarly, Christiansen and Chater (2016) found that comprehension question accuracy is generally low in temporarily ambiguous sentences and is accompanied by a higher degree of re-reading of text. However, re-reading in their study also did not predict comprehension accuracy. They suggest that re-reading plays a "confirmatory" role, not a revisionary one. Along the same lines, other authors have suggested that wrap-up costs are "a checking process to make sure that processing is complete before the memory representation of lower-level information is discarded at the end of a sentence" (Stowe et al., 2018, p. 233). Similarly, Hirotani et al. (2006) consider the Dwell Time Hypothesis, that readers dwell to do a secondary check for infelicities, akin to "pausing before [leaving] the house" (p. 426), due to a reluctance to move on from material that will shortly be displaced from working memory. They further suggest that wrap-up costs are therefore not modulated by the difficulty or amount of remaining linguistic work associated with the current clause.

Taken together, these suggestions lead to the conclusion that wrap-up effects are at least sometimes a product of non-linguistic processes that involve reflecting on one's understanding following the initial processing of a segment, and that perhaps the effort expended on such reflection is a product of the difficulty of the task conditions. While we find promise in the intuition behind such proposals, many are underspecified or empirically untestable. Nevertheless, the general idea points to a role for metacognition/comprehension in wrap-up costs – the process of introspecting on cognitive processes (*metacognitive monitoring*) in order to effectively self-regulate memory and comprehension (*metacognitive control*) (Rhodes, 2016).

Consistent with this view, another line of related work suggests that readers allocate attention to boundary positions in a strategic and self-regulated manner (S. Andrews & Veldre, 2021; Stine-Morrow et al., 2010). This explains the tendency to dwell on some boundaries but not others. Stine-Morrow et al. (2010) proposed that readers adopt an individual-level resource allocation policy that leads to a *pay-now-or-pay-*

later tradeoff: that the probability of dwelling on a particular boundary is a function of dwell time on previous boundaries, and that different individuals vary in their tendency to pay early vs. late. For example, Stine-Morrow et al. (2010) found that older readers were more likely to pay a high wrap-up cost early, but show smaller wrap-up times at downstream boundaries. On the other hand, younger adults were more likely to pay late. Stine-Morrow et al. suggested that this asymmetry by age is a means of offsetting working memory load for older adults, who typically have lower working memory capacity. S. Andrews and Veldre (2021) expand on Stine-Morrow et al.'s work by showing that the pay-now-or-pay-later tradeoff is associated with individual differences in reading proficiency. They used a number of reading and vocabulary tests to create a combined index of reading proficiency (see the paper for more detail), then investigated reading patterns for high vs. low proficiency readers, finding the same asymmetry reported by Stine-Morrow et al.. Specifically, lower proficiency readers were more likely to dwell earlier and longer but not dwell again later, whereas higher proficiency readers showed smaller wrap-up costs at earlier boundaries but longer reading times overall at late boundaries irrespective of punctuation.

❧ Wrap-up as a hybrid process

It's very likely that the sources reviewed above jointly contribute to wrap-up effects. Sentence comprehension is a complex process that requires recruiting memory, integrating various sorts of linguistic information, making decisions about how to allocate processing resources, and reflecting on one's understanding. In line with this view, some researchers have proposed hybrid accounts of wrap-up. For example, the *buffer-integrate-purge* model (Haberlandt & Graesser, 1989; Magliano, Graesser, Eymard, Haberlandt, & Gholson, 1993) suggests that surface level representations are actively held in memory until clause boundaries, where they are converted into more durable, higher-level linguistic representations that can be integrated into the discourse representation, then removed from the working memory buffer in order to free up resources for further processing. Stine-Morrow et al. (2010)'s account adopts these assumptions too, but further suggests that individual differences interact with readers' decision of *when* to integrate.

While the Context-Sensitive Encoding hypothesis does not posit a working memory buffer, it provides a foundation that supports a similar view of wrap-up effects. This relies on three mechanistic assumptions: (i) that prosodic boundaries influence the probability of shifting the context vector, (ii) that resulting encoding contexts can be reinstated at later boundaries, (iii) that reinstating a context takes up time and resources, and the pause in other incremental processes afforded by reaching an implicit prosodic boundary allows for this, and (iv) that different individuals will opt to reinstate previous contexts at different times, depending on their working memory capacity and other factors. In addition, the additional processing time afforded by prosodic boundaries may provide participants with the chance to perform discourse updating processes, like changing the status of a particular segment or establishing a relationship between clauses, given sufficient evidence in the discourse context.

4.3.1.1 Context-Sensitive Memory at Boundaries

As a starting point, the current experiment aimed to investigate the reading time pattern and memory processes proposed in Experiment 5, using the same constructions. Experiment 6 thus used self-paced reading followed by a recognition memory task. We contrasted prosodically integrated vs. isolated AWA-coordination, as in (69), investigating readers' wrap-up times and temporal order memory for the adjective-noun pairs within each coordinate (e.g., *scary butchers* vs. *lazy butchers*, and *lazy bakers* vs. *scary bakers*).

(69) a. ISO: Ramona believed that the scary butchers, as well as the lazy bakers,

were lying about the ingredients in the recipe.

b. INT: Ramona believed that the scary butchers as well as the lazy bakers were lying about the ingredients in the recipe.

We sought to investigate three related areas: (i) wrap-up effects at AWA boundaries, (ii) the sentence-final memory representation of each coordinate, and (iii) how wrap-up correlates with memory for each coordinate. We expected to verify the wrap-up differences at AWA boundaries discussed in §4.2.5. We predicted that under Context-Sensitive Encoding, we should observe (i) better memory for both the first and second coordinates in prosodically isolated constructions compared to integrated ones and (ii) a positive correlation between wrap-up time at each boundary and memory for the contents of the immediately preceding prosodic phrase. Because Hirotani et al. (2006) observe wrap-up costs only at the right edge of ARCs, we alternatively predicted that memory for both coordinates may correlate with time at this boundary alone. In order to investigate Stine-Morrow et al.'s (2010) pay-now-or-pay-later strategy, correlation with time at the sentence-final region was also measured, although we did not have specific predictions about this region.

Under Reinstantiation, we predicted that (i) memory for an isolated first coordinate should be better than an integrated first coordinate, but that (ii) in addition, memory for the first coordinate should be better than for the second coordinate, because by hypothesis, it is activated three times (during initial encoding, at the right boundary of the second coordinate, and at the main verb), whereas the second coordinate is active only once (during initial encoding) in isolated structures, and once or twice in integrated structures (during initial encoding and at the main verb, some proportion of the time). Therefore, we predicted (i) that within the second coordinate, memory for the integrated structure should be better than the isolated structure, because it is more likely to benefit from reactivation, and (ii) that time spent at the right AWA boundary should correlate with memory for the first coordinate.

A secondary analysis sought to determine whether our data replicated Stine-Morrow et al.'s pay-now-or-pay-later tradeoff for high and low performance on the recognition memory task. We had reason to expect that this tradeoff may interact with memory for each coordinate, so we further investigated this split in the context of the recognition memory results and correlational analyses.

4.3.2 Method

4.3.2.1 Participants

113 participants from Prolific were recruited to take part in the experiment. Participation restrictions were identical to those used for Experiments 2-3. Each experimental session took approximately 50-60 minutes, and participants were compensated $12/hr. 7 participants were excluded due to less than 80% accuracy on the distractor math task. The analysis included data from 106 participants.

4.3.2.2 Materials

56 item sets crossing Structure (INT, ISO), Position (COORD-1, COORD-2), and Match (MATCH, MISMATCH) were constructed, as in Table 4.12. The Structure factor varied whether the sentence participants read contained prosodically isolated or integrated AWA-coordination. Recognition probes targeted adjective-noun relationships in order to assess memory for each coordinate without targeting only the phrase-final words, because these are also the words that we hypothesized participants would spend more time on in ISO conditions. Participants were asked to provide Related/Unrelated judgments on the critical adjective-noun pairs. MATCH conditions probed memory for previously read adjective-noun associations, whereas MISMATCH conditions presented participants with nouns paired with an incorrect adjective (the one that appeared in

the non-target coordinate). Word pairs were always presented in backwards order (e.g., an adjective-noun sequence like *scary butchers* was presented as *butchers-scary* in the recognition probe), so that participants could not rely on forward sequential information or a surface match in memory alone to provide their judgment. The Position factor varied whether the recognition probe asked about association with N1 (COORD1) or with N2 (COORD2).

Structure	Sentence	regions of interest
ISO	Ramona believed that the scary butchers, as well as the lazy bakers, were lying about the ingredients in the recipe.	
INT	Ramona believed that the scary butchers as well as the lazy bakers were lying about the ingredients in the recipe.	

(a) Sentences by Structure for Experiment 6. ***Key***: R1: *butchers(,)*, R2: *bakers(,)*, R3: *lying*, R4: *recipe*.

	MATCH	MISMATCH
COORD1	butchers-scary	butchers-lazy
COORD2	bakers-lazy	bakers-scary

(b) Recognition probes by condition for Experiment 6.

Table 4.12: Sample item set for Experiment 6.

Because wrap-up times are sensitive to low-level properties of the words at boundaries, we controlled for word frequency and length following S. Andrews and Veldre (2021). Word frequency statistics were measured using OpenLexicon (Pallier, New, & Bourgin, 2019). We controlled for `logSUBTLWF` and contextual diversity, such that N1 and N2 were within 0.2 units of each other on both measures. In addition, N1, N2, and the sentence-final word were within +/-2 characters of one another in length. The prenominal adjectives were controlled (relative to each other) in the same manner. The sentence-final word was matched on length but not frequency. Because we sought to investigate memory for each of the coordinates, we ensured that the first (A1) and second (A2) adjectives were semantically unrelated to one another. We matched A1 and

A2 on semantic relatedness using `word2vec` (Foltz, n.d.; Mikolov et al., 2013) scores as a proxy. Only adjectives with `word2vec` scores < 0.35 were included in the critical positions. The average A1-A2 `word2vec` score was 0.19.

Experimental items were presented across 8 Latin-squared lists along with 54 fillers. Fillers contained a mixture of appositive and restrictive relative clauses, *not only...but also* and *...and also...* constructions, and conjoined transitive clauses. The subject and object positions sometimes contained simple noun phrases, and sometimes contained NP of NP constructions. Participants were asked to judge the relationship between pairs of words in different regions and syntactic positions of the sentences. Probe types included adjective-noun, subject-object, numeral-noun, and NP of NP relationships, so that participants could not simply rely on the temporal order of adjacent words, but had to remember the syntactic relationships between words as well. Selected examples are given in (70)-(71).

(70) Mimi gifted handmade jewelry to her grandkids, but Rich only gifted a few pairs of boring socks.

   a. MATCH: jewelry-Mimi

   b. MISMATCH: socks-Mimi

(71) Janine invited the son of the nomad, and also the aide of the teacher, even though she didn't know either of them.

   a. MATCH: teacher-aide

   b. MISMATCH: teacher-son

4.3.2.3 Procedure

The experiment utilized self-paced reading followed by a recognition memory task. Participants read sentences using a word-by-word, non-cumulative, moving window self-paced reading paradigm. They then responded to a distractor arithmetic problem by typing in their response. Arithmetic problems involved addition or subtraction of two randomly generated numbers between 0-100, where the answer was always a positive value. After the distractor phase, they were asked to provide a judgment of whether a noun-adjective probe matched the association between an adjective-noun relation in the sentence they read (*Were these words related in the sentence you read?*). Judgments were provided on a 6-point scale from -3 (*no, definitely not*) to 3 (*yes, definitely*) so that we could construct empirical ROC curves from participants responses (see §3.2.2.2).

4.3.3 Results

Participants with < 80% accuracy on the math distractor task were excluded, leaving 106 participants in the analysis. In addition, trials on which participants were +/- 2 standard deviations from the mean math reaction time were excluded. Self-paced reading trials with reaction times < 150 ms or > 5000 ms were also excluded. Average judgment accuracy was 77% (78% for experimental items and 75% for fillers).

4.3.3.1 Self-Paced Reading Results

Word-by-word self-paced reading latencies by Structure are plotted in Figure 4.5. Throughout this section, the regions of interest are labeled as in Table 4.12a, where *butchers(,)* is labeled as R1, *bakers(,)* as R2, *lying* as R3, and *recipe* as R4.

At the first region of interest (*butchers(,)*), ISO conditions displayed slower latencies than INT. Two words down (at *well*), ISO conditions showed evidence of a compensatory speed-up, as has been reported in other work investigating post-comma reading times (Hirotani et al., 2006). The the second region of interest (*bakers(,)*), ISO conditions were once again slower than INT. At the verb (*lying*), reaction times show a crossover pattern, where INT latencies are slower than ISO ones. Although this mirrors

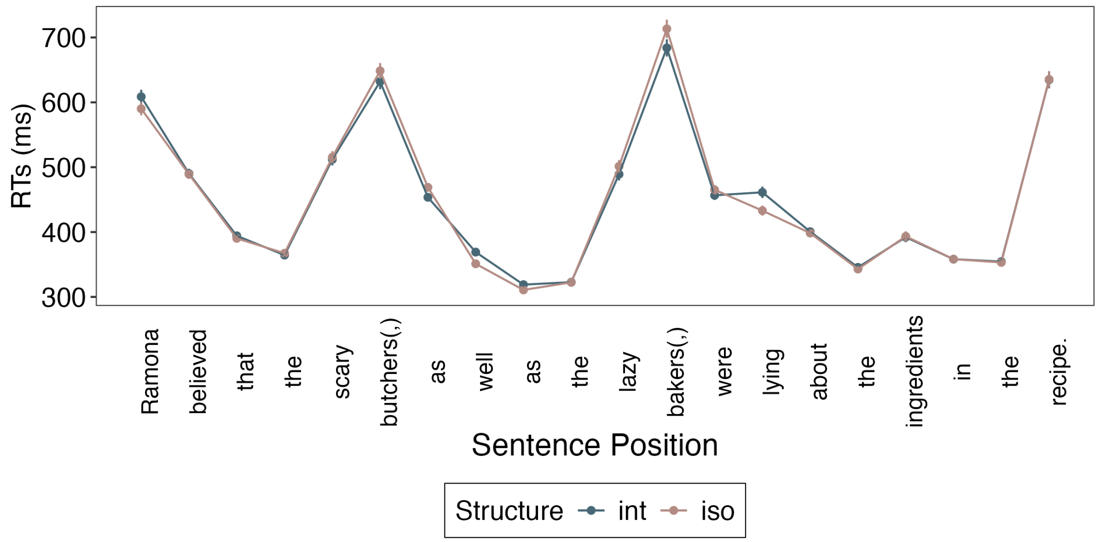


Figure 4.5: Word-by-word SPR latencies by Structure for Experiment 6.

the compensatory speed-up following the edge of the first coordinate, the difference in latencies at the verb is larger. We attribute this to the bypassing effect we reported in Experiment 5: that retrieval interference is reduced for ISO structures compared to INT. Finally, the sentence-final region displayed no difference between structures.

Because we had reason to expect a larger wrap-up penalty at the right edge of the second coordinate, we investigated the interaction between Structure (ISO, INT) and Region (R1, R2). Bayesian linear mixed effects models[38] were fit to response latencies at the final region of the first coordinate (*butchers(,)*) and the final region of the second coordinate (*bakers(,)*)) using `brms` (Bürkner, 2017). Predictors were sum-coded, with the INT and R1 conditions mapped to negative values. Model results are in Table 4.13. $\hat{R}$-values and posterior predictive checks indicated model convergence for all models reported in the current section.

The model results revealed a main effect of Structure such that ISO > INT ($\hat{\beta}$ = 0.02, CrI = [0.0001, 0.04]), and a main effect of Region such that times to R1 < R2 ($\hat{\beta}$ = 0.03,

[38]`brm(logRT ~ Structure*Region + (1 + Structure*Region | Subject) + (1 + Structure*Region | Item))`

Effect	$\hat{\beta}$	95% CrI
Structure	**0.02**	**(0.0001,0.04)**
Region	**0.03**	**(0.008,0.04)**
Structure*Region	-0.0003	(-0.04,0.03)

Table 4.13: Bayesian linear mixed-effects models fit to log RTs at regions R1 and R2 of Experiment 6.

	Verb (*lying*)	
Effect	$\hat{\beta}$	95% CrI
Structure	**0.04**	**(0.01,0.06)**

Table 4.14: Bayesian linear mixed-effects models fit to log RTs at the verb region (R3) of Experiment 6.

CrI = [0.008, 0.04]). The interaction between Structure and Region was not credible ($\hat{\beta}$ = -0.0003, CrI = [-0.04, 0.03]).

In addition, we fit a `brms` model[39] to response latencies at R3, the verb region. The Structure predictor for this model was treatment coded, with INT as the reference level. These results are in Table 4.14. This model revealed a credible difference such that latencies on the verb were faster for ISO conditions compared to INT ones ($\hat{\beta}$ = 0.04, CrI = [0.01, 0.06]).

Finally, we investigated the relationship between RTs at R1, R2, and R4 in order to investigate the pay-now-or-pay-later tradeoff by judgment accuracy. We split participants into two groups – low vs. high sensitivity – based on those participants whose mean d' value was below the median value vs. above the median value, respectively. These results are plotted in Figure 4.6.

The high sensitivity group generally displayed longer reaction times than the low sensitivity group. SPR latencies at each region also qualitatively displayed a pay-now-or-pay-later tradeoff analogous to the pattern reported by Stine-Morrow et al. (2010)

[39] `brm(logRT ~ Structure + (1 + Structure | Subject) + (1 + Structure | Item))`

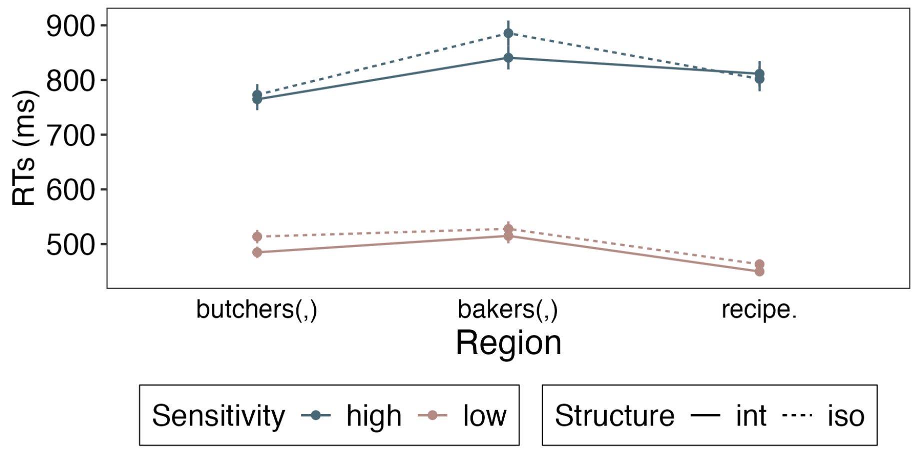


Figure 4.6: SPR latencies by condition for low vs. high sensitivity participants in Experiment 6.

and S. Andrews and Veldre (2021). The low sensitivity group more more likely to dwell early (at R1, *butchers,*) but not at R2 (*bakers,*) or R4 (*recipe*). The high sensitivity group did not exhibit a wrap-up cost at R1, but dwelled longer on ISO structures at R2, and did not dwell at R4.

To investigate the validity of this pattern, Bayesian linear mixed effects models[40] were fit to latencies at each boundary region (R1, R2, and R4). Predictors were sum-coded, with INT and LOW conditions mapped to negative values. Results (in Table 4.15) suggested that the qualitative pattern described above was not reliable, as the credible interval for the Structure x Sensitivity interaction crosses 0 for all three regions. The only credible result was a main effect of Sensitivity such that the high sensitivity group exhibited longer RTs overall than the low sensitivity group. This effect emerged across all three regions. Though the interaction between Structure and Sensitivity was not reliable, we discuss this further in the discussion, because the recognition memory results and time-sensitivity correlations diverged by group in notable ways.

[40] `brm(logRT ~ Structure*Sensitivity + (1 | Subject) + (1 + Structure | Item))`

Effect	R1 (*butchers(,)*) $\hat{\beta}$	95% CrI	R2 (*bakers(,)*) $\hat{\beta}$	95% CrI	R4 (*recipe.*) $\hat{\beta}$	95% CrI
Structure	0.02	(-0.001,0.05)	0.02	(-0.01,0.05)	0.01	(-0.01,0.04)
Sensitivity	**0.36**	**(0.17,0.56)**	**0.42**	**(0.19,0.65)**	**0.38**	**(0.19,0.57)**
Struc*Sens	-0.01	(-0.06,0.04)	0.01	(-0.04,0.06)	0.01	(-0.03,0.06)

Table 4.15: Pay-now-or-pay-later analysis for Experiment 6, with Bayesian linear mixed-effects models fit to log RTs at R1, R2, and R4.

4.3.3.2 Signal Detection Theory Results

The judgment data for the recognition memory task were subjected to a unequal variance Signal Detection Theory analysis (Hautus et al., 2021), using the pROC package in R (Robin et al., 2011) with the same procedure as described in §3.2.2.2 for Experiment 1. Empirical ROCs are plotted in Figure 4.7, with all conditions scaled against their respective MISMATCH conditions. We calculated sensitivity, in terms of d_a and area-under-the-curve (AUC) values, for each condition. Statistical comparison of AUC values was performed with pROC, using 2000 bootstrap replicates. pROC model results are provided in Table 4.16.

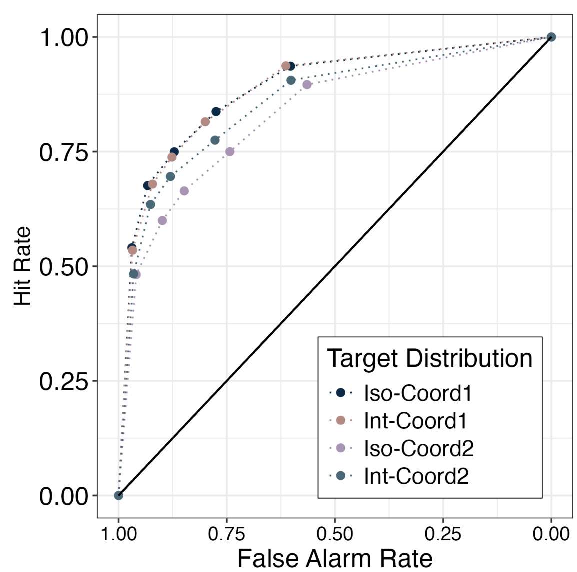


Figure 4.7: Empirical ROC curves by condition for Experiment 6.

	d_a	AUC	2.5%	97.5%
ISO-COORD1	1.7	0.89	0.87	0.9
INT-COORD1	1.7	0.89	0.87	0.9
	D_{boot} = 0.08		p = 0.9	
ISO-COORD2	1.3	0.83	0.81	0.85
INT-COORD2	1.5	0.86	0.84	0.88
	D_{boot} = -1.9		p = 0.05	
COORD1	1.7	0.89	0.87	0.9
COORD2	1.2	0.84	0.83	0.86
	D_{boot} = 4.4		p < 0.001	

Table 4.16: `pROC` model results for Experiment 6.

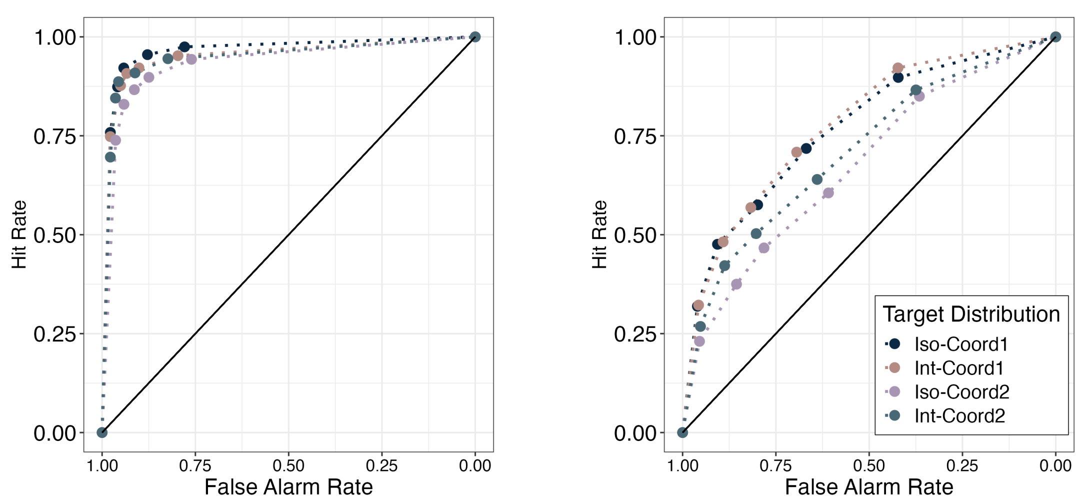
There was no significant difference in sensitivity between ISO-COORD1 (d_a = 1.7) and INT-COORD1 (d_a = 1.7) conditions (p = 0.9). INT-COORD2 (d_a = 1.5) conditions displayed significantly higher sensitivity than ISO-COORD2 (d_a = 1.3) conditions (p = 0.05). In addition, sensitivity to COORD1 (d_a = 1.7) conditions was greater than sensitivity to COORD2 (d_a = 1.2; $p < 0.001$).

4.3.3.3 Time-Sensitivity Correlations

We calculated time-sensitivity correlations between by-participant wrap-up cost (ISO RTs - INT RTs) and by-participant difference in sensitivity (ISO d' - INT d') per Position. Because high- vs. low-sensitivity groups of participants displayed different reading time patterns, we measured correlation coefficients for high- vs. low-sensitivity groups in the current analysis. The low-sensitivity group included participants whose overall sensitivity level was less than or equal to the median by-participant d'; the high-sensitivity group included participants whose sensitivity was above the median. Spearman correlation coefficients were calculated using the `cor.test()` function in R. Results are plotted in Figure 4.9, along with correlation coefficients. The R1-COORD2 panel is absent because we did not expect times at the edge of the first coordinate to influence sensitivity to the second coordinate's content.

The high-sensitivity group did not display a statistically significant relationship between wrap-up time and sensitivity at R1 or R2. In contrast, the low-sensitivity group exhibited a significant positive correlation between time and sensitivity at R1 (R = 0.3, p = 0.03) and a marginally significant positive correlation between time and sensitivity for COORD2 at R2 (0.26, p = 0.07). In addition, the high-sensitivity group displayed a marginally significant negative correlation for COORD2 at R4 (R = -0.24, p = 0.09).

This pattern of results suggests that the predictions of CSE obtained only for the low-sensitivity group. On the face of it, this result is somewhat puzzling. In order to further explore high- vs. low-sensitivity differences, we compared empirical ROCs for each sensitivity group. These are plotted in Figure 4.8. There were not enough observations to subject these subsets to statistical tests, so we discuss trends in the data here instead.



(a) High-sensitivity ROCs. **(b)** Low-sensitivity ROCs.

Figure 4.8: Empirical ROCs by high- vs. low-sensitivity groups for Experiment 6.

Visual inspection of these plots suggests that the pattern of results in the aggregate dataset was primarily driven by the low-sensitivity group. The difference in sensitivity to COORD1 vs. COORD2 is much larger for the low-sensitivity group. However, both groups displayed lower sensitivity to ISO-COORD2. The average d' value for the high-sensitivity group was 3.44; this is effectively at ceiling (Hautus et al., 2021). In contrast, average d' value for the low-sensitivity group was 1.1. We suspect that the much longer RTs displayed by the high-sensitivity group reflected a much more elaborate reading strategy that led to near-perfect performance. From this perspective, it is no wonder that time-sensitivity correlations in the low-sensitivity group were more interpretable. It also suggests that high- vs. low-sensitivity groups employed different reading strategies. We discuss this pattern further in the discussion.

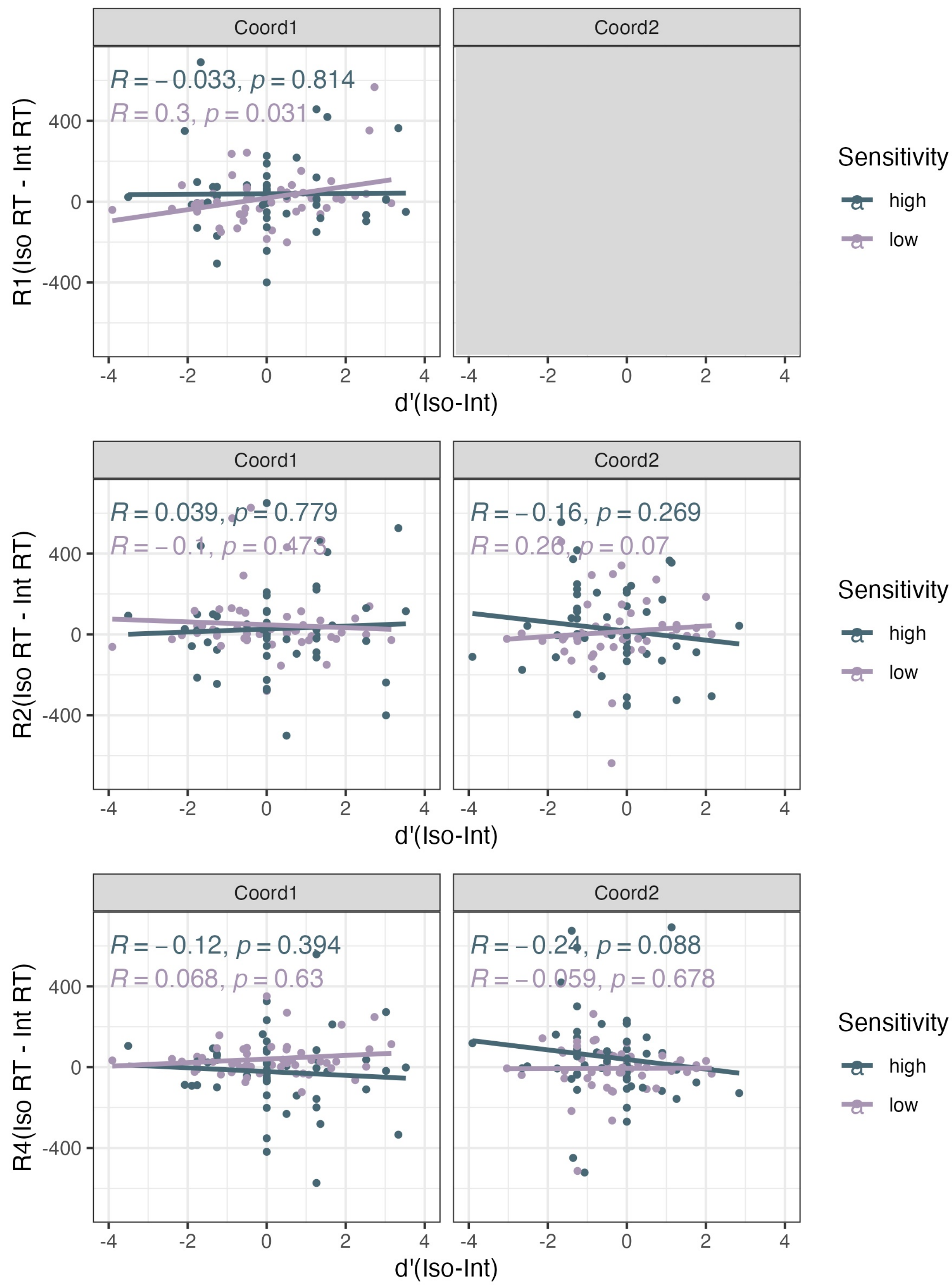


Figure 4.9: Time-sensitivity correlations by Position and high- vs. low-sensitivity groups for Experiment 6.

4.3.4 Discussion

Beginning with a discussion of the self-paced reading results, we observed the presence of wrap-up effects such that RTs to ISO structures were longer than INT at both the first and second boundary positions, in line with Context-Sensitive Encoding and Reinstantiation. We initially expected to find a larger wrap-up cost at the second boundary (R2), per Hirotani et al.'s (2006) reading time data on appositive relative clauses. Reinstantiation also predicts a larger cost at the second boundary than the first, because this is the hypothesized locus of contextual reinstatement. Numerically, our results trended in this direction, as the difference between ISO vs. INT reaction times was smaller at R1 than at R2. However, the interaction between Structure and Region was not credible. We think the particular task we utilized may have influenced readers' tendency to engage in wrap-up processes more often. Recall Haberlandt et al.'s finding that recall tasks yielded much larger wrap-up costs than comprehension and unstructured reading tasks. The current experiment involved recognition memory, whereas Hirotani et al.'s study utilized eye-tracking while reading only. Therefore, it is reasonable to assume that the memory-oriented nature of our task encouraged readers to wrap up at both boundaries. This is further supported by the fact that in Experiment 5, we saw a signature of wrap-up at the right, but not the left, AWA-coordinate boundary.

In line with Reinstantiation, we also replicated our finding from Experiment 5 that prosodically isolated AWA-coordinates engender less retrieval interference than their integrated counterparts. This was evidenced by longer reading times for INT structures at the verb (R3). Notably, this result emphasizes the role of unambiguous cues to prosodic boundaries in triggering shifts in the contextual representation during encoding. Even for discourse-identical comparisons, the presence of prosodic boundaries is sufficient to prompt bypassing. It also rules out the alternative focus-based explanation considered for Experiment 5: that RTs at the verb to prosodically isolated AWA were faster than for ordinary coordination because AWA-is a focus-sensitive structure, where N1 and N2 are encoded more deeply, and thus more accessible in memory. We assume, following Hulsey (2008), that both the ISO and INT structures in the current experiment were focus-identical. Therefore, the current results cannot be attributed to differences in focus-structure or discourse independence.

The SDT results were partially consistent with Reinstantiation. Specifically, we found a significant difference between sensitivity to ISO vs. INT structures in the second coordinate, such that sensitivity to ISO-COORD2 was worse. Recall the predictions of Reinstantiation: that in isolated structures, the context associated with the first coordinate should be reinstated twice, once at the right edge of coordinate two, and once again upon successful retrieval of N1 at the verb. In contrast, the context associated with the second coordinate in ISO structures is hypothesized to be active only during initial encoding. These predictions align with domain-general work on memory, which has established that opportunities for retrieval practice strengthen the representation of items in long-term memory, leading to better retention. Work on the role of temporal context and retrieval practice comes to a similar conclusion (Jang & Huber, 2008; Lehman, Smith, & Karpicke, 2014; Sahakyan & Hendricks, 2012). For example, the *list-before-last* paradigm shows that being cued to retrieve a prior list context strengthens the representation of items in that list, and reduces activation/interference from an intervening list (the most recently encoded one). Specifically, the context associated with the second list does not benefit from retrieval practice, and so the representation of the second list is ultimately degraded relative to the first one. If the prosodic boundary following the second coordinate in ISO structures serves as a cue to reinstate the context associated with the first coordinate, coordinate one should ultimately be better retained in memory than ISO-COORD2. In contrast, INT-COORD2 remains accessible along with INT-COORD1, because both the first and second coordinate occupy the same encoding context in INT structures, and thus the first coordinate is not reinstated, but either COORD1 or COORD2 may benefit from retrieval at the verb.

Under CSE, which underlies Reinstantiation, we should have also observed a difference between ISO vs. INT structures in the first coordinate, due to the benefit of contextual partitions. Furthermore, under Reinstantiation, more opportunities for contextual reinstatement of the first coordinate in isolated structures also should have resulted in a benefit for ISO-COORD1. One possibility is that the content of the first coordinate benefits from a primacy effect, causing INT structures to be remembered equally well. Beyond this, we don't presently have an explanation for why memory for ISO-COORD1 did not show a benefit[41].

Despite this, the time-sensitivity correlations supported the role of CSE for both coordinates. The high-sensitivity group was at ceiling in the recognition memory task, so we do not expect correlations for this group to be meaningful. Within the low-sensitivity group, we found that longer wrap-up times related to a larger benefit for memory in ISO compared to INT conditions. This effect emerged at both AWA boundaries, but not the sentence-final region. This suggests that the distinction in encoding context imposed by prosodic boundaries does facilitate better temporal order memory within each coordinate. This pattern of results aligns directly with domain-general investigations of temporal order memory across group boundaries as well (Pu et al., 2022a). Under Reinstantiation, we predicted that wrap-up times at R2 should positively correlate with sensitivity to COORD1 content, if the first coordinate is reinstated at this point. We did not observe any evidence of this. At R2, a positive time-sensitivity correlation was observed only for COORD2 content, not COORD1. Therefore,

[41]It may be that participants adopted the strategy of consulting their memory for COORD1 in order to make judgments about both COORD1 and COORD2. This would render ISO/INT-COORD1 equally memorable, but lack of contextual similarity between COORD1 and COORD2 should make accessing ISO-COORD2 relatively more difficult.

the correlational analysis provided support for Context-Sensitive Encoding, but not Reinstantiation.

One reason for this may be that the process of contextual reinstatement is not reflected by the amount of time spent dwelling on a boundary. §4.3.1 established that a number of processes, both linguistic and non-linguistic, occur at boundary positions, and that individual readers may use time spent dwelling in distinct ways. This makes it conceptually difficult, and further undesirable, to reason that the absence of a correlation implicating a relationship between wrap-up time and a particular process suggests that the process does not occur at all. Because Reinstantiation was validated by the self-paced reading results, and partially validated by the recognition memory results, we argue that it is relevant to the processing of ARCs and AWA-coordination. We may have failed to observe its effect in time-sensitivity correlations because time spent dwelling on boundaries primarily reflects other processes. Additionally, if the processing load associated with contextual reinstatement (retrieving a contextual state) is akin to retrieving an item from memory, as the TCM suggests, this may not be an elaborate operation that takes a significant amount of time.

Interestingly, in the sentence-final region, we found tentative support for a negative relationship between wrap-up time and sensitivity within the high-sensitivity group only. This may point to the conclusion that some distinct process applies at sentence-final boundaries, whereby COORD2 content becomes less available. Alternatively, this could be consistent with the predictions of Reinstantiation. If, by the end of the sentence, the context of the first coordinate has been reinstated multiple times, it's possible that this relates to a relative decrement in memory for coordinate two, under the assumption that increasing the activation of one coordinate's representation takes away activation from the other. However, the fact that this tracks along with the amount of time spent on R4 is unexpected. We don't at present have an explanation for this. Note that this particular result should be interpreted with caution, because the negative correlation at this region was only marginally significant, and it emerged for the group whose recognition memory performance was at ceiling.

Finally, we entertain a competing explanation. At first glance, the data may seem consistent with a hypothesis like Compression, which would predict no difference in sensitivity to the contents of the first coordinate, but predicts worse memory for isolated coordinate two content. As stated in §3.1.3, Compression is a process that is argued to apply due to the background/subordinate linguistic status of a particular unit. AWA-coordination, then, does not provide the appropriate linguistic context to observe its effect. But for the sake of argumentation, let's suppose that some more general notion of backgrounding could drive Compression of particular segments in this case. Even if this were true, we could still rule out a Compression-based explanation because it is inconsistent with the SDT results of Experiment 1: there, we saw that memory for the syntactic structure of ARCs was better than that of RRCs. Furthermore, a Compression-based account is inconsistent with the correlational data we report here. We generally found that longer wrap-up times correlate with greater sensitivity, whereas Compression might have predicted the opposite (i.e., that longer times correlate with worse sensitivity), because the content of the second coordinate is inhibited at this point in time. Finally, the identical reading time profile of ARCs and isolated AWA-coordination gives us reason to believe that there is not a difference between these constructions that would lead to Compression in one case and CSE/Reinstantiation in the other. In sum, we do not take the results to be consistent with Compression.

4.4 General Discussion

The current chapter uncovered several pieces of evidence in favor of the Reinstantiation hypothesis, the basic claim of which is that the contents of an initial encoding context may be reinstated after the processing of a subsequent context (at its right boundary), which has the effect of reducing interference from the second context. The original framing of Reinstantiation in (72), which was proposed to account for bypassing of appositives, now requires some refining. After summarizing the main findings of this chapter, I put forward a revised definition.

(72) **Reinstantiation** (original): The contents of an earlier encoding context (= main clause) can be reinstated using positional cues in order to add additional material to a previously incomplete segment (c.f. Sederberg et al., 2011), such that the initial and final segments of the main clause in sentences containing ARCs are encoded as contextually similar to one another because they have a syntactic relationship.

Experiment 5 of this chapter determined that AWA-coordinates are by-passed, just like ARCs. Experiment 6 confirmed that Reinstantiation is a process that occurs at unambiguous prosodic boundaries (i.e., those overtly marked by orthographic cues in written text). Therefore, we concluded that the bypassing of AWA-coordinates in Experiment 5 cannot be attributed to their focus structure or to differential discourse status. Furthermore, Experiment 6 suggested that time spent dwelling at comma-marked words at the left and right boundaries of AWA-coordinates led to a benefit for memory for each coordinate. Together, these results point to the suggestion that the parser utilizes extra processing time at sentence-medial intonational phrase boundaries to partition linguistic segments in the memory representation of a sentence and to reinstate previous encoding contexts at boundary positions immediately prior to when their contents must be utilized for subsequent processing operations (e.g., establishing a subject-verb dependency). For ARCs and AWA-coordination, this results in relatively higher accessibility for the contents of the initial segment at the time entering the final segment. Therefore, items encoded in the initial segment are also more likely to be accessed during retrieval later on. Additionally, more opportunities to reinstate the initial encoding context in isolated structures leads to better retention of that content relative to the medial encoding context. We suggested that this effect is related to Jang and Huber's (2008) finding that retrieval practice for an initial list context leads to worse retention of a medial list context.

Finally, Experiment 4 attempted and failed to search for evidence of Context-Sensitive Retrieval in NOBA structures with a single sentence-medial prosodic boundary, but with focus and discourse characteristics otherwise similar to AWA-coordination. CSR assumes that retrieving an item should reinstate that item's encoding context, and so the amount of potential interference should be reduced in segmented structures. Our results did not support this view. We did, however, find evidence that could have been consistent with Visibility: that the resolution of anaphora across a prosodic boundary resulted in a decision time penalty. This view could be reconciled with Reinstantiation. If the ease of retrieval in ARCs and AWA-coordinate structures comes from a prosodic cue to pre-active a prior contextual state before encountering the site of a dependency, the work of contextual reinstatement is done ahead of time. However, if accessing an item outside the current encoding context comes with the cost of reinstating that context, per CSE-Visibility, we should expect to see that cost emerge at the site of the dependency. This suggests a difference between sentences with one medial boundary, which should only induce a context shift, and those with two medial boundaries, which should encourage contextual reinstatement (specifically in particular syntactic positions). A significant issue with this logic is that we do not observe a cost for accessing items across a sentence-final ARC boundary during retrieval, suggesting that item-level retrieval cues prevail without issue in these cases. I leave this issue open for now, but discuss it further in Chapter 5. Therefore, we consider our alternate explanation – that the cross-boundary anaphora penalty in Experiment 4 reflected a violation of focus structural expectations – to be a more likely explanation.

Taken together, the results support the overarching conclusion that sentence-medial prosodic boundaries encourage reinstatement of earlier contexts in memory, specifically in advance of an anticipated subject-verb dependency. However, I note that AWA-coordination is different than ARCs in that both coordinates have a syntactic relationship with the following verb, whereas in ARCs, only the head noun of the relative clause has a syntactic relationship with the main verb[42]. For ARCs, this offered a linguistic explanation for why the contents of the initial segment of the main clause may be privileged over the contents of the appositive prior to encountering the main verb. It's less clear that this explanation can transfer straightforwardly to AWA-coordination, where N1 and N2 can both agree with the verb and must both be retrieved at the verb in order for complete coordinate resolution to take place. Therefore, I suggest that the prosodic boundary positions of ARCs and AWA-coordination, along with the dynamics of reinstating contexts, can explain why the bypassing effect obtains for both constructions. On this basis, I offer the revised definition in (73).

(73) **Reinstantiation** (revised): The contents of a prior encoding context (= prosodic phrase) can be reinstated using positional cues in order to refresh its representation in memory and integrate its contents with upcoming material (c.f. Sederberg et al., 2011) in anticipation of an upcoming syntactic dependency. Contextual reinstatement has two effects: it drives *isolation* between the current

[42] As far as agreement within a cue-based retrieval framework goes.

(medial) and upcoming contexts, reducing interference from the medial context, and it leads to greater contextual similarity between the initial and upcoming contexts (c.f. Jang & Huber, 2008).

This results in a contextual representation as in (74), where each encoding context is given a label (A, B, or C), purely for the sake of expository convenience, not because I assume context-level features of this sort. Here, context C is associated with its own contextual features, but also has features in common with context A, which was reinstated immediately prior to the encoding of context C. Under this approach, context B remains distinct although it also has a syntactic relationship with contexts A/C because the disruption of reinstating context A after the processing of context B leads to greater contextual distinction between B and C. This suggestion is analogous to the *list isolation* proposal by Jang and Huber (2008): that in a situation where participants learn three lists (L1, L2, and L3) and are cued to retrieve L1 following the encoding of L2, this leads to greater contextual distinctiveness between L2 and L3, but greater similarity between L1 and L3. This framing argues that the bypassing effect (the fact that context B contributes less to interference) is driven primarily by prosodic segmentation and the order of memory operations at boundaries, not the syntactic relationships between segments.

(74) CSE + Reinstantiation (revised)

1 2 3

$\boxed{\text{The}_{A}\ \text{chefs}_{A},}$ $\boxed{\text{as}_{B}\ \text{well}_{B}\ \text{as}_{B}\ \text{the}_{B}\ \text{butchers}_{B},}$ $\boxed{\text{have}_{A/C}\ \text{prepared}_{A/C}\ \text{a}_{A/C}\ \text{recipe}_{A/C}}$

1 : *chefs,* → shift from $\boxed{\text{context A}}$ to $\boxed{\text{context B}}$

2 : *butchers,* → shift from $\boxed{\text{context B}}$ to $\boxed{\text{context C}}$, reinstate $\boxed{\text{context A}}$

3 : *have* → encode $\boxed{\text{context C}}$ along with features of $\boxed{\text{context A}}$;

retrieve *chefs*, because $\boxed{\text{context A}}$ is more accessible

In sum, (74) claims that reinstating context A immediately prior to encoding context

C has two effects: (i) it effectively isolates context B, rendering it more contextually distinct, and (ii) at the same time, it decreases contextual distinctiveness between context A and C, such that the contents of C retains some of its own contextual features, but also bears contextual similarity with context A.

This is not to say that syntactic relationships between segments cannot influence their contextual encodings. In ARCs for example, the head of the relative clause contained within context A has a syntactic relationship with an argument position in context B. Similarly in AWA-coordination, coordinate one of context A and coordinate two of context B are syntactically connected. In principle, then, it is possible that syntactic relationships could influence contextual similarity between segments, too. In fact, this could potentially explain the fact that linguistic content appears to interact in the Main $\rightarrow$ ARC direction but not the ARC $\rightarrow$ Main direction (Dillon et al., 2017; S. Kim & Xiang, 2022). Suppose that the contents of context B in (74) are encoded with context features that are specific to context B but also similar to context A. This could be possible if forward-looking syntactic dependencies, like filler-gap processing, affect context encodings. Suppose also that the reinstatement of context A prior to context C drives isolation of context B. This could explain why main and ARC content appear to interact prior to the right boundary of an ARC, but not afterwards. But on the other hand, if context B were also encoded with features of context A, it would be unclear why contextual reinstatement of A would not bring along context B as well (see discussion of Sederberg et al., 2011, in §2.3.3).

For these reasons, the role of prosodic boundaries offers a more explanatory account of bypassing effects, but I acknowledge that any temporal context mechanism applied to the domain of language must bring greater specificity to the question of how contextual representations evolve in the presence of multiple interconnected linguistic relationships. It could be that the nature of context shifts are such that they are sensitive to initial groupings, which will often align with the first-pass prosody of a sentence when prosodic boundaries are unambiguously marked, and that these groupings can sometimes serve to facilitate access particular linguistic segments in a more targeted manner. But, this does not require that the emergent encoding contexts are strictly linguistic in nature. They could be initial partitions in memory that cue-based retrieval is sensitive to and that have consequences for linguistic processing. But, their formation need not be driven in a top-down manner by linguistic parsing processes themselves. Alternatively, the temporal context mechanism may interface only with certain levels of linguistic structure (e.g., the prosody, but not the syntax). I leave these issues for future work, but to be clear, it is also important to draw a distinction between how encoding contexts are formed and how they are used. While their formation may be driven by more general temporal groupings, I've argued here that they can be strategically used during language processing.

Chapter 5

Reanalysis across Contexts

So far, this dissertation has advanced the idea that prosodic boundaries partition sentences into distinct encoding contexts in memory via a temporal context mechanism, per the Context-Sensitive Encoding hypothesis. In the previous chapter, I established that these contexts can be manipulated during online comprehension such that a previous contextual state can be reactivated at particular sentence-medial prosodic boundary positions, per Reinstantiation, and upcoming content can be encoded as contextually similar to the reinstated context. However, a fundamental assumption of temporal context models is that item-level retrieval drives the evolution of the contextual representation, because retrieval of an item reinstates that item's context, and the retrieved context along with new contextual features is associated with upcoming items. In Chapters 3-4 we attempted to extend this premise to the domain of sentence comprehension through the Context-Sensitive Retrieval (CSR) hypothesis. We explored evidence for CSR using noun phrase ellipsis spanning prosodic/discourse segment boundaries, to little avail. Cue-based retrieval of items in segmented structures did not reliably result in facilitated dependency resolution in ARC, NOBA, or AWA structures.

However, item-level dependencies in sentence processing may be the wrong place to search for an effect of CSR. Contextual reinstatement of this sort is hypothesized to occur following successful retrieval of an item. That is, the retrieval of an item "brings along" features of its encoding context, so to speak. Therefore, the dependencies we investigated may weight item-level features so heavily that any effect of contextual interference is rendered more or less inert, especially because successful comprehension requires moving forward with parsing the remainder of the sentence once an item-to-item dependency has been established successfully. In other words, the timescale and dynamics of parsing may not lend themselves to observing an effect of CSR, because sentence processing requires returning to the next item in the current encoding context shortly after accessing the context of the targeted item.

Against this background, we hypothesized that the effect of Context-Sensitive Retrieval should be more evident for sentence processing operations where the retrieved context matters for the dependency at hand. Chapter 5 investigates reanalysis, a syntactic process that has downstream interpretive consequences, and requires the comprehender to "look beyond" a retrieved item to its local syntactic structure in order to facilitate successful comprehension. Reanalysis partially relies on cue-based retrieval (Van Dyke & Lewis, 2003), but also requires revision of initially encoded syntactic structure. Thus, we sought to understand how retrieval of an item in a segmented structure interacts with a comprehender's ability to reactivate the syntactic structure built at or near the time of initial encoding. To probe this, in Experiments 7 and 8 we investigated syntactic reanalysis of NP/S ambiguities in sentences containing prosodically isolated AWA-coordination with multiple loci of reanalysis that spanned encoding context boundaries.

The remainder of this chapter reviews other work on the interpretive consequences of prosodic phrasing (§5.1), introduces NP/S ambiguities, the type of reanalysis utilized in Experiments 7-8 (§5.2), then outlines the predictions of relevant hypotheses, including CSR (§5.3), before turning to the details of the experiments (§5.4-5.5).

5.1 Interpretive Domains

Prosodic boundaries constrain interpretation. As discussed in §2.1.2.2, they have an immediate effect on syntactic parsing, as they can rescue the parser from pursuing an incorrect parse before disambiguating input has been encountered. Consequently, prosodic boundaries in dispreferred early closure positions that align with the syntactic parse of the sentence, as in (1), are facilitated. Conversely, in the wrong position like in (2), they can further lure the parser into pursing the incorrect analysis, only to create greater processing difficulty later on (Kjelgaard & Speer, 1999). In implicit prosody, garden path sentences that require revisions of both syntactic and prosodic boundary positions are more difficult for comprehenders to recover from than those that require syntactic revisions alone (Bader, 1998). These findings point to a central role for prosodic boundaries in grouping syntactic content.

(1) When Roger leaves, the house is dark.

(2) *When Roger leaves the house, is dark.

Prosodic boundaries can also bound interpretations in a way that leads to ungrammaticality. For example, Frazier et al. (2014) investigated the effect of prosodic boundaries on cases of *local coherence*, in which comprehenders tend to favor locally licit but globally ungrammatical syntactic parses over globally grammatical ones. An example from Tabor, Galantucci, and Richardson (2004) is given in (3), where the locally coherent parse (*the player tossed the frisbee*) is bracketed.

(3) The coach smiled at [the player tossed the frisbee].

(3) exemplifies a reduced relative clause garden path construction (*The coach smiled at the player (who was) tossed the frisbee*), which is independently known to lead to

processing difficulty. After reading such sentences, there is a tendency for participants to incorrectly respond *yes* to the question *Did the player toss the frisbee?* Frazier et al. (2014) investigate the role of prosodic boundaries in facilitating these ungrammatical interpretations. They hypothesized that if an illicit root clause (*the player tossed the frisbee*) is self-contained within its own prosodic phrase, listeners should be more likely to fall for the incorrect interpretation. They manipulated the position of a prosodic boundary such that the illicit root clause was phrased alone, or it was phrased together with the main clause verb, as in (4a).

(4) a. ($_\varphi$The kindergarteners at the school) ($_\varphi$liked the little girl brought a toy).

b. ($_\varphi$The kindergarteners at the school liked) ($_\varphi$the little girl brought a toy).

In line with their predictions, Frazier et al. (2014) found that locally coherent interpretations are more likely given the phrasing in (4b), where the illicit root clause "stands alone" in its own prosodic phrase.

Other work has claimed that intonational phrase boundaries create interpretive domains, whose boundaries delineate points where outstanding semantic/pragmatic processes are completed. This idea is captured by Schafer's (1997) Interpretive Domains Hypothesis. While the definition of IDH itself is somewhat underspecified, it has a clear connection to the work on wrap-up effects discussed in Chapter 4, which proposes conceptual integration at prosodic boundaries.

(5) **Interpretive Domains Hypothesis** (IDH) (Schafer, 1997)

An intonational phrase boundary defines a point at which the processor performs any as yet outstanding semantic/pragmatic evaluation and integration of material within the intonational phrase.

Schafer tests the IDH using lexical ambiguities across two listening experiments. An example item set from one of these experiments, which investigated the interpretation of homonyms such as *glasses*, is in Table 5.1. Schafer assumed the following sequence of processes for these items. Following work on the processing of homonymy (Duffy, Morris, & Rayner, 1988; Rayner & Frazier, 1989), encountering the lexically ambiguous word *glasses* should prompt immediate semantic commitment to the preferred meaning (*spectacles*), and because the context of the first phrase is consistent with both meanings, the frequency of each meaning should dictate which interpretation listeners pursued first. The IDH forces "outstanding" semantic evaluation at intonational phrase boundaries, and so it predicts that listeners should more deeply process the meaning of the initial clause when it is followed by an intonational phrase boundary. Schafer suggests that deeper processing in this case amounts to generating more inferences about the preceding material. In contrast, a phonological phrase boundary[43] in the same position should not encourage deeper semantic/pragmatic evaluation. In the HI-FREQ conditions where the main clause meaning is consistent with the preferred meaning of *glasses* (*Stacey wore them anyway*), interpretation proceeds without issue. In the LO-FREQ conditions, however, earlier commitment to the preferred meaning of *glasses* followed by disambiguation to the dispreferred, low-frequency meaning (*a drinking container*) should lead to processing difficulty, specifically in the condition with a prior intonational phrase boundary, under the assumption that reanalysis across an ιP boundary is more costly.

Schafer measured reaction times to each condition in a sentence-final "makes sense" task and found a main effect of Meaning, such that RTs to LO-FREQ conditions were longer than those to HI-FREQ conditions, and an interaction between Meaning and Boundary Type, such that the disambiguation cost was greater for low-frequency meanings in intonational phrase conditions. This was taken as support for the IDH,

[43]The syntax-prosody correspondence introduced in §2.1.2.1 would suggest that an ι-boundary should occur in this position, not a φ-boundary, because intonational phrases are roughly clause/proposition-sized. However, Schafer (1997) assumes optionality between intonational and phonological phrases in this syntactic position.

Meaning	φ-BOUND	ι-BOUND
HI-FREQ	($_\varphi$Although the glasses were ugly) ($_\varphi$Stacey wore them anyway)	($_\iota$Although the glasses were ugly) ($_\iota$Stacey wore them anyway)
LO-FREQ	($_\varphi$Although the glasses were ugly) ($_\varphi$they held a lot of juice)	($_\iota$Although the glasses were ugly) ($_\iota$they held a lot of juice)

Table 5.1: Example item set from Schafer (1997), Experiment 4.

because reanalysis to the dispreferred lexical meaning in a neutral context was more difficult across an intervening ιP boundary. This experimental task did not constitute a perfect test of the IDH, as it relied on sentence-final reaction times measured several words down from the first cue to disambiguating input, a rather coarse measure.

A follow-up cross-modal naming study (Schafer & Speer, 1998) corroborated the predictions of the IDH using an online methodology. There, when presented with a context that was strongly biased towards the dispreferred meaning of a homonym (*($_{\iota/\varphi}$Since the **anchor** is more attractive now) ($_{\iota/\varphi}$...)*), cross-modal naming times to visual probes that related to the preferred (BOAT) or dispreferred (NEWS) meaning were compared. Schafer and Speer found that RTs were facilitated for the dispreferred meaning only across ιP boundaries. The authors took this to mean that ιP boundaries encouraged deeper commitment to the dispreferred meaning, thus facilitating access to that meaning later on, whereas φP boundaries did not have this effect. Together, the results support the general idea that the strength of a prosodic boundary modulates the ease of reanalysis for pre-boundary content, such that revision across stronger boundaries (ιPs) is more costly.

Such a view also explains why in reading of ambiguous garden path sentences, recovery from the incorrect parse is often difficult. Consider the NP/Z garden path in (6), which was introduced in §2.1.2.2.

(6) While Anna dressed the baby spit up on the bed.

Recall that misinterpretations in these cases tend to linger, in that even if readers display an online reanalysis cost at the verb (*spit up*), in sentence-final comprehension tasks they maintain that both the misanalysis (*Anna dressed the baby*) and the revised interpretation (*the baby spit up on the bed*) are correct (Ferreira et al., 2001). This tendency has been termed the *semantic persistence effect*; I discuss this further in the following section. Suppose that the initially preferred parse here encourages comprehenders to posit an intonational phrase boundary following *the baby* in the implicit prosody, and this in turn encourages integration of the initial clause (*While Anna dressed the baby*) with the discourse representation (one possible outstanding semantic/pragmatic process). Then upon encountering the disambiguating input, complete revision of the misanalysis will prove difficult if it remains intact at multiple levels of representation (i.e., the syntax and the discourse).

In general, such findings suggest that once the parser reaches an intonational phrase boundary, revision of previous content is difficult, and in some cases, largely unsuccessful. Under the *buffer-integrate-purge* view of wrap-up effects (Haberlandt & Graesser, 1989; Magliano et al., 1993), it may be that once the content of a prosodic phrase has been integrated into the discourse representation, and by hypothesis converted into a more durable, long-lasting form, that content is also more difficult to access or revise. This is an alternate way to frame the types of discourse processes that occur at ιP boundaries. Instead of appealing to to-be-completed semantic/pragmatic processing, it could simply be that it takes time to repackage content into a gist-level (Potter & Lombardi, 1990) format that can be integrated into the discourse representation.

Interestingly, this consequence of intonational phrase boundaries (difficulty with reanalysis) does not seem to apply to appositive relative clauses (Dillon et al., 2018). I discuss this in the following section, after introducing some basic information about the processing of NP/S ambiguities, the construction of interest in Experiments 7-8.

5.2 NP/S Ambiguities

NP/S garden path sentences involve a temporary ambiguity between direct object (NP) and sentential complement (S) parses at the point of encountering the underlined NP *the butcher* in (7).

(7) Ramona believed <u>the butcher</u>...

a. NP-Comp parse: Ramona [$_{VP}$ believed the butcher]...

b. S-Comp parse: Ramona believed (that) [$_{S}$ the butcher ...]

(8) Ramona believed **the butcher** <u>was</u>...

At *the butcher*, there is an initial preference to pursue the NP parse in (7a). Given additional disambiguating input (underlined in (8)), participants display a slowdown in reading and reaction times compared to the same region in unambiguous constructions, where the sentential complementizer *that* is overt (*Ramona believed that the butcher...*). The penalty at the embedded clause verb suggests that the parser engages in some type of revisionary process to correct the initial misanalysis. Crucially, reanalysis of the previous parse must occur in order to continue integrating incoming input into the currently-being-built syntactic structure and to arrive at the correct sentence-final interpretation.

The mechanism underlying reanalysis is contested in the sentence processing literature. Broadly, theories of reanalysis (and syntactic parsing more generally) fall into two camps. On the one hand, serial models assume that only one parse can be pursued at a time. In NP/S garden paths, this is the initially preferred NP parse. Then when the parser encounters disambiguating input, it must make "edits" to the previously built structure before continuing to integrate incoming input. Reanalysis of NP/S garden paths has been argued to require four such edits: (i) de-linking of the NP *the butcher* from the matrix verb *believe*, (ii) switching the lexical frame of *believe* from a transitive verb to one that can take a sentential complement, (iii) re-attaching *the butcher* as the subject of the upcoming clause, and (iv) and attaching the upcoming clause itself to the matrix verb (Lewis, 1998; Van Dyke & Lewis, 2003). On the other hand, parallel models assume that the parser can simultaneously construct multiple syntactic analyses for an ambiguous string, and that these parses are weighted by likelihood or frequency. Under these models, the source of the slowdown at the disambiguating region is not due to syntactic editing, but rather, re-ranking of the already constructed parses based on the evidence at hand. In the current chapter, we adopt a serial view of reanalysis.

One account of reanalysis, proposed by Van Dyke and Lewis (2003), argues that the processes required during reanalysis can be explained under a cue-based retrieval framework as follows. First, the authors assume that the relative frequencies of the NP vs. S frames of the matrix verb drive initial attachment decisions, after which the not-chosen frame (i.e., the S-Comp frame of *forgot* in (9)) begins to decay in memory. Then, once disambiguating input (*was*) has been encountered, Van Dyke and Lewis' proposal is that the retrieval probe on that input prompts retrieval of a clausal attachment site using the features {+V, +S-COMP}. This search returns the matrix verb *forgot*, which could have had such an attachment site, if the other parse were pursued. Finally, following retrieval of the S-Comp frame of *forgot*, an additional probe on *was* triggers retrieval of a subject in the standard manner, using the cues +NP, +NOMINATIVE, and +SINGULAR. This search returns the NP *student*. These retrieval operations are depicted in (9), recreated from Van Dyke and Lewis (2003).

(9) The secretary **forgot** the student who was waiting for the exam **was**...

   a. Retrieval of clausal attachment site

+V,S-COMP ?V,S-COMP

forgot the student who was waiting for the exam was

b. Retrieval of subject

+N,NOM,SG ?N,NOM,SG

forgot the student who was waiting for the exam was

Key: match partial match

I refer the reader to Van Dyke and Lewis (2003) for a more extensive discussion, but importantly, this suggests that the effects of attachment and reanalysis are distinguishable. Here, I make some additional assumptions about how these processes may be related to one another. One notable feature of their proposal is that the required clausal attachment site at *forgot* in (9a) is only a potential one, but is not fully realized because the S-Comp frame of *forgot* is never selected. While the alternative S-Comp frame receives some activation at the time of encoding *forgot*, it is ultimately overridden by the decision to pursue the preferred NP-Comp parse. This is why Van Dyke and Lewis assume that the S-Comp lexical frame decays. It's not entirely essential to assume retrieval of a potential attachment site must occur, and in fact, recent research has called into question the claim that attachment sites can be retrieved (Ben-Meir, Van Handel, & Wagers, 2019). As such, we assume that retrieval is driven by item-specific cues, rather than potential attachment sites, and offer some modifications to Van Dyke and Lewis's (2003) proposal. The proposed timecourse of how retrieval of subjects might guide reanalysis is outlined in (10).

(10) Tasks of the parser in NP/S ambiguities, based on (9):

a. Initially attach *the student who…* as the object of *forgot*

b. Disambiguating V *was* triggers retrieval of a potential subject, *student*

c. Reactivation and search of local syntactic structure from retrieved NP

d. De-linking of *student* from the matrix verb *forgot*, only if the frame of *forgot* is successfully replaced

Specifically, the process described in (10) proceeds as follows. First, the NP-Comp parse is initially pursued (10a). Upon encountering the disambiguating input, the verb initially triggers subject retrieval, not attachment site retrieval (10b). Given successful retrieval of the subject *the student*, I assume that attachment to the verb proceeds in a manner analogous to J. D. Fodor and Inoue's (1998) *Attach Anyway* principle, though I don't commit to whether this requires the parser to represent a globally ungrammatical resulting structure in memory[44]. Then, the syntactic structure local to the retrieved subject is reactivated and searched for the locus of the initial misanalysis. For the example described in (10c), this only requires searching the immediately dominating node in the syntactic tree. As stated in (10d), if this search successfully results in retrieval of the alternate frame of *forgot*, then *the student* is successfully de-linked from *forgot*, and the incorrect parse is fully re-analyzed.

Importantly, step (10d) leaves room for reanalysis to fail, even if attachment succeeds. This is a desirable consequence, as *semantic persistence* effects are well-established for garden path sentences, as alluded to previously. For both NP/S and NP/Z garden paths, it is common for the initially pursued, incorrect analysis to linger. In support of this point, Sturt (2007) found that NP/S garden paths like in (11) exhibit a reading time penalty at the disambiguating region *was*, but also exhibit a penalty at the sentence-final region *impossible to reach*. This region contributes a meaning that is inconsistent with the initially preferred parse, because it cannot be that *the explorers found the South Pole* and that it was *impossible to reach*.

[44] One could imagine a situation where two locally coherent sub-trees exist simultaneously for some period of time, until the reanalysis process is fully resolved.

(11) The explorers found the South Pole **was** impossible to **reach**.

One explanation that has been put forth for this observation suggests that syntactic parsing is sometimes "good enough"; that is, the parser may sometimes build underspecified or incomplete syntactic structures, resulting in long-lasting misinterpretations (Ferreira et al., 2001). However, Slattery, Sturt, Christianson, Yoshida, and Ferreira (2013) report the results of two eye-tracking studies that provide evidence against this claim: their studies show that lingering misinterpretations are not due to a failure to arrive at the structure associated with the correct interpretation, but rather a failure to clean up the remnants of the incorrect parse. Adopting an approach where constructing the correct analysis proceeds via ordinary cue-based retrieval of a subject, but where revising the incorrect analysis is dependent on accessibility of (i) the to-be-revised structure and (ii) the alternate frame of the verb in memory neatly captures Slattery et al.'s observation.

The processes in (10) interact in an important way with prosodic boundaries. I discuss this in detail shortly, but first return to a point raised in the previous section. Recall that prosodic boundaries play a significant role in mediating recovery from garden path sentences. The previous section established that positing a prosodic boundary in an incorrect position, "wrapping up" the interpretation of the preceding content, and moving on to parse subsequent input has a detrimental effect on the parser's ability to recover from syntactic misanalysis.

Notably for current purposes, the relationship between prosodic boundary positions and syntactic reanalysis plays out differently for appositive relative clauses. In an acceptability judgment study on NP/S ambiguities spanning ARCs vs. RRCs, Dillon et al. (2018) found that reanalysis cost for NP/S garden paths spanning ARCs is no greater than those spanning RRCs, as in Table 5.2. The authors take this as evidence that switching between not-at-issue and at-issue segments does not strengthen

Structure	Ambiguity: -AMBIG = ∅, +AMBIG = *that*
RRC	Becca found {∅,that} the security guard who Anne argued is a trained cop fell asleep on duty.
ARC	Becca found {∅,that} the security guard, who Anne argued is a trained cop, fell asleep on duty.

Table 5.2: Example item set from Dillon et al. (2018), Experiment 3.

the comprehenders' commitment to the initially preferred NP-complement parse, or render main clause content less accessible (contra the predictions of the Visibility and Interpretive Domains hypotheses).

Dillon et al. (2018) suggest that the matrix clause verb *found* is equally accessible for reanalysis regardless of whether an ARC or an RRC intervenes between it and the disambiguating region. Under a view where main clause content is reinstated at the right edge of an ARC, as proposed in Chapters 3-4, this is somewhat puzzling. In theory, Reinstantiation of the main clause immediately prior to encountering the disambiguating input in ARCs should make it easier for the parser to access the locus of misanalysis, and thus lead to a smaller reanalysis cost for ARCs compared to RRCs. Conversely, one might expect the presence of intonational phrase boundaries in ARCs to cause the processor to commit more deeply to the incorrect NP-complement parse at each boundary, and to further reinforce that commitment during contextual reinstatement of the initial encoding context immediately prior to re-entering the main clause, where the disambiguating region resides. This state of affairs would be consistent with the general role of prosodic boundaries on interpretation, and should predict more difficult with reanalysis in ARC structures. Thus, there are two ways that Reinstantiation could affect reanalysis, but neither of these predictions are borne out in Dillon et al.'s results. It's possible that both processes may be active to some extent – that attachment would be easier for segmented structures due to Reinstantiation, but reanalysis is more difficult because the presence of prosodic boundaries, along with contextual reinstatement, reinforces the incorrect interpretation.

A limitation of Dillon et al.'s study is that it only assesses the ultimate acceptability of NP/S ambiguities, not the accessibility of relevant pieces of syntactic structure during online comprehension, or the interpretation that readers ultimately arrive at. Even if attachment is equally easy across ARCs and RRCs, revision of the misanalysis may not be. To explore this in more depth, we turn to an investigation of how the prosodic boundaries in AWA-coordinate constructions affect reanalysis of content in previous prosodic phrases.

5.3 Context-Bounded Reanalysis

The remainder of this chapter reports the results of two experiments. The first one (Experiment 7) uses B-SPR to compare reading times on the disambiguating region for the two constructions in (12a)-(12b) compared to ordinary coordination (12c) in order to investigate Context-Sensitive Retrieval in a situation that encourages reactivation of surrounding syntactic structure. The second one (Experiment 8) uses ordinary SPR followed by a sentence-final comprehension question task to compare the degree of semantic persistence for the misanalysis associated with each coordinand.

(12) a. N1-ISO: Ramona believed the **butchers**, as well as the chef, **were** preparing an innovative menu for the restaurant.

b. N2-ISO: Ramona believed the butcher, as well as the **chefs**, **were** preparing an innovative menu for the restaurant.

c. BOTH-INT:Ramona believed the **butchers** and the **chefs were** preparing an innovative menu for the restaurant.

Note that encountering the disambiguating verb phrase in (12a)-(12b) requires revision of *two* misanalyses: both the first coordinate (*Ramona believed the butcher(s)*), and the second coordinate (*Ramona believed the chef(s)*), under the assumption that focus-sensitive ellipsis is reconstructed during the processing of the second coordinate (see §4.1.1 and §4.2.1 for details about the structure of focus-sensitive coordination). Following (10), suppose that the retrieval cues on the disambiguating verb guide retrieval of the subject(s), and consequently also affect reanalysis. Specifically, the degree of match between the features on the disambiguating verb and each coordinand should modulate the success of constructing the correct analysis. When the retrieval probe matches the contents of one coordinate but not the other in an isolated structure, Context-Sensitive Retrieval assumes retrieval of the better-matching noun phrase and reactivation of only the syntactic structure contained within its encoding context. Thus, the presence of prosodic boundaries delineating encoding contexts should lead to reduced accessibility of the content of the non-matching coordinate during reanalysis, because the retrieved context of the matching coordinate will not be a good cue to the contextual features of the other coordinate, and the item-level cues on the retrieval probe also won't be as good of a match to the non-matching coordinate. In contrast, both coordinands in the integrated structure should be equally accessible, because they both match the retrieval probe, and they match each other more closely in contextual features. Therefore, the cost of reanalysis should be greater in the integrated structure. These predictions are schematized in (13).

(13) Predictions under Context-Sensitive Retrieval

a. N1-ISO < BOTH-INT

Ramona believed the **butchers** | <~~Ramona believed~~> the chef | **were**...

b. N2-ISO < BOTH-INT

Ramona believed the butcher | <~~Ramona believed~~> **the chefs** | **were**...

In (13a), the verb better matches the features of the first coordinand, whereas in (13b), it better matches the features of the second one. If the early stages of reanalysis are guided by the degree of match between the verb and the potential subject(s), and if reactivation of syntactic structure in these situations is constrained by the targeted noun's encoding context, per CSR, this should differentially modulate accessibility for each of the coordinates in (13), and also differentially affect reanalysis. In other words, once the targeted noun has been accessed and its context reactivated, it may be difficult for the parser to access the other coordinate if the retrieved context does not provide a clear retrieval cue to other encoding contexts.

Therefore, CSR predicts easier access to the coordinates of both ISO structures compared to the INT baseline. This prediction is tested in Experiment 7, and its results are confirmed. Crucially, it also predicts more successful reanalysis of the first coordinate in the ambiguous N1-ISO condition, and more successful reanalysis of the second coordinate in the ambiguous N2-ISO condition. These predictions are tested in Experiment 8, which compares rates of semantic persistence for each coordinate.

Work on the interpretive effects of prosodic boundaries and the source of clause-final wrap-up raise another possibility: that the presence of an intonational phrase boundary encourages the parser to strengthen commitment to the parse currently being built, perhaps because it encourages integration of the current segment into the discourse representation, which transforms that segment into a less malleable format in memory. This should predict more difficulty revising content across prosodic boundaries. This may be further exacerbated for ambiguous N1-ISO whose encoding context, by hypothesis, is reinstated at the second prosodic boundary. We term this possibility Interpretive Domains (+ Reinstantiation), schematized in (14).

Crucially, this hypothesis predicts that attachment should proceed without significant issue, if the retrieval mechanism can recruit a syntactic level of representation.

But reanalysis, on the other hand, should be more costly for structures where the parser engages in intermediate wrap-up and earlier portions of the sentence are sent off to be integrated with a higher-level, discourse representation. We assume that syntactic memory survives this process, as the experiments up until this point have shown that for the most part, item-to-item dependencies are not hindered when they span prosodic boundaries (contra Visibility).

(14) Predictions under Interpretive Domains (+ Reinstantiation)

a. N1-ISO > BOTH-INT

Ramona believed the **butchers** <~~Ramona believed~~> the chef **were...**

b. (N1-ISO >) N2-ISO > BOTH-INT

Ramona believed the butcher <~~Ramona believed~~> **the chefs** **were...**

Finally, we consider the possibility that Reinstantiation facilitates reanalysis of the first coordinate, if the retrieval cues on the disambiguating region match N1. In this case, we would expect reanalysis of N1-ISO to be less costly than both N2-ISO and BOTH-INT, as in (15). Reinstantiation alone predicts no particular difference between N2-ISO and BOTH-INT.

(15) Predictions under Reinstantiation

a. N1-ISO < BOTH-INT

Ramona believed the **butchers** <~~Ramona believed~~> the chef **were...**

b. (N1-ISO <) N2-ISO = BOTH-INT

Ramona believed the butcher <~~Ramona believed~~> **the chefs** **were...**

The predictions of the three hypotheses are summarized in Table 5.3.

Ramona believed...

	Context-Sensitive Retrieval	
Structural Configuration	Coordinate 1	Coordinate 2
the rumors, as well as the tabloid, **were**	✔ (< C2)	✗
the rumor, as well as **the tabloids**, **were**	✗	✔ (< C1)
	Reinstantiation	
Structural Configuration	Coordinate 1	Coordinate 2
the rumors, as well as the tabloid, **were**	✔ (< C2)	✗
the rumor, as well as **the tabloids**, **were**	✗	✗
	Interpretive Domains + Reinstantiation	
Structural Configuration	Coordinate 1	Coordinate 2
the rumors, as well as the tabloid, **were**	✗ (>C2)	✗
the rumor, as well as **the tabloids**, **were**	✗ (>C2)	✗

...fabricated by the tabloid.

Table 5.3: Predicted reanalysis cost under CSR and two versions of Reinstantiation for Experiment 8. ***Key***: Structural configuration {✔ = does, ✗ = does not} facilitate reanalysis.

5.4 Experiment 7

Experiment 7 first establishes the reanalysis profile of the constructions in (12) using B-SPR. We also aimed to determine whether the content of the first and second coordinates in AWA-constructions are equally accessible, like in the case of ARCs.

5.4.1 Method

5.4.1.1 Participants

60 participants were recruited via Prolific with the same restrictions used in the previous experiments. Each experimental session took approximately 45-60 minutes, and participants were compensated $12 per hour for their participation. 7 participants were excluded from the analysis due to the fact that they did not attain at least 70%

Match	Ambiguity (+AMBIG = ∅, -AMBIG = *that*) **critical** | spillover
BOTH-INT	Ramona believed {∅,that} the rumors and the articles rather embarrassingly **were** fabricated by the tabloid over the summer.
N1-ISO	Ramona believed {∅,that} the rumors, as well as the article, rather embarrassingly **were** fabricated by the tabloid over the summer.
N2-ISO	Ramona believed {∅,that} the rumor, as well as the articles, rather embarrassingly **were** fabricated by the tabloid over the summer.

Table 5.4: Example item set from Experiment 7.

comprehension question accuracy on either the experimental items or the fillers. An additional 5 participants recruited from the University of California, Santa Cruz Linguistics Department Subject Pool were used as replacements and were granted course credit for their participation. The final analysis included data from 58 participants.

5.4.1.2 Materials

60 item sets, each consisting of 6 conditions, were constructed. The design crossed two factors: Ambiguity (+AMBIG, -AMBIG) with Match (BOTH-INT, N1-ISO, N2-ISO). The Ambiguity factor varied whether the complementizer *that* was overt. The Match factor varied whether the verb matched N1, N2, or both in number features. The baseline BOTH-INT conditions contained ordinary coordination, whereas the N1/N2-ISO conditions contained prosodically isolated AWA-coordination. Table 5.4 contains a sample item set.

In the N1-ISO conditions, the verb better matched N1 in number features. In the N2-ISO conditions, the verb better matched N2. Finally, the BOTH-INT condition acted as a baseline, involving ordinary coordination and a configuration where both coordinates matched the verb in plural number features.

As the experiment investigated NP/S ambiguities, matrix verbs were always ambiguous in whether they could take NP- or S-complements; all verbs were taken from Van Dyke and Lewis (2003) or Sturt (2007), who report previous investigations of NP/S ambiguities that used verbs with an NP-complement preference. Each of these verbs was used no more than 3 times across the full item set. The reanalysis trigger was always an auxiliary that bore plural number features. In addition, the following verb (*fabricated*) was selected to be incompatible with the NP-complement parse, following the design in Sturt (2007). In our items, we attempted to ensure that the incompatibility was localized to this verb as much as possible. In Sturt's items, on the other hand, the implausibility of the NP-Comp parse was sometimes not made apparent until the end of the sentence (e.g. *The explorers found the South Pole was impossible to* ***reach***). We made this design decision in order to bias readers against semantic persistence effects. Finally, these sentences also contained a pre-critical adverbial region, as in the Experiment 5, in order to mitigate potential reaction time facilitation following comma-marked words. Because these adverbs were part of the temporarily ambiguous region, they were constructed to be consistent with either the NP-Comp (*Ramona believed the rumors and the articles rather embarrassingly*) or S-Comp parses (*The rumors and the articles rather embarrassingly were fabricated*). We controlled for this so that adverbials could not act as an early reanalysis trigger prior to the critical region. However, they were not normed prior to conducting the experiment, and could have been interpreted parenthetically, which may have led to the insertion of an implicit prosodic boundary prior to the auxiliary[45]. The results suggest that if this occurred, it did not collapse differences between the conditions. In any case, we did away with this region in Experiment 8.

Experimental items were counter-balanced across 6 Latin-squared lists and presented along with 44 fillers in randomized order. Filler sentences contained other focus-sensitive coordination constructions with sentence-medial prosodic boundaries and NP/Z garden path sentences in order to disguise the ambiguous experimental items.

[45]Thanks to Maziar Toosarvandani for this observation.

5.4.1.3 Procedure

The procedure was identical to that of Experiment 5: participants read sentences using moving window, word-by-word B-SPR and answered a comprehension question following half of the experimental trials and half of the filler trials.

5.4.2 Results

Average comprehension question accuracy was 73% (77% for the experimental items and 70% for the fillers). We attribute the relatively low comprehension accuracy for the fillers to the presence of NP/Z garden path sentences, but note also that the filler comprehension accuracy in the current experiment was comparable to Experiment 5. Accuracy on the comprehension questions for the experimental items was lower than in Experiment 5, however; this was likely due to the fact that 2/3 of the questions targeted the temporarily ambiguous region of the NP/S constructions (sometimes the question asked about N1, sometimes about N2, and sometimes about other region of the sentences).

Reading measures (first-pass, go-past, and total times) were calculated using the `em2` package (Logacev & Vasishth, 2013) in R. We fit Bayesian linear mixed-effects models[46] using the `brms` package (Bürkner, 2017) to the critical and spillover regions for each of the three reading measures using informative priors from Paape and Vasishth (2021). The Ambiguity predictor was sum-coded, with the -AMBIG condition mapped to a negative value, and the Match predictor was treatment coded, with the BOTH-INT condition set to the reference level. $\hat{R}$ values and posterior predictive checks at the critical and spillover regions indicated model convergence. First-pass, go-past, and total reading times are plotted in Figure 5.1. Table 5.5 contains regression weights

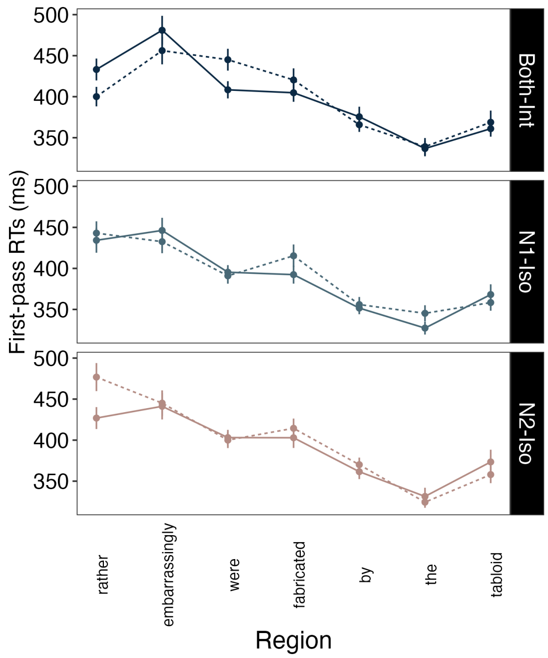
[46] `brm(logRT ~ Ambiguity*Match + (1 + Ambiguity*Match | Subject) + (1 + Ambiguity*Match | Item))`

and 95% credible intervals from models fit to log RTs on the critical and spillover regions.

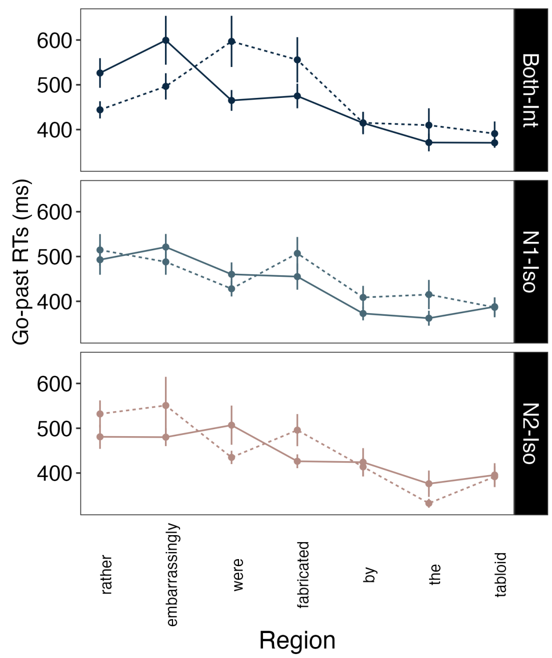
	Critical (*were*)		Spillover (*fabricated*)	
First-Pass RTs	$\hat{\beta}$	CrI	$\hat{\beta}$	CrI
Ambiguity	**0.05**	**(0.01,0.10)**	0.02	(-0.03,0.06)
Match-N1	**-0.05**	**(-0.08,-0.02)**	-0.02	(-0.06,0.01)
Match-N2	**-0.03**	**(-0.07,-0.00)**	-0.01	(-0.04,0.02)
Amb*Match-N1	**-0.06**	**(-0.12,-0.00)**	0.03	(-0.03,0.09)
Amb*Match-N2	**-0.07**	**(-0.14,-0.01)**	0.01	(-0.05,0.08)
Go-Past RTs	$\hat{\beta}$	CrI	$\hat{\beta}$	CrI
Ambiguity	**0.08**	**(0.02,0.14)**	0.03	(-0.03,0.1)
Match-N1	**-0.07**	**(-0.11,-0.03)**	-0.03	(-0.08,0.01)
Match-N2	**-0.05**	**(-0.1,-0.00)**	-0.03	(-0.07,0.01)
Amb*Match-N1	**-0.1**	**(-0.18,-0.02)**	0.02	(-0.06,0.11)
Amb*Match-N2	**-0.12**	**(-0.21,-0.03)**	0.01	(-0.07,0.09)
Total RTs	$\hat{\beta}$	CrI	$\hat{\beta}$	CrI
Ambiguity	**0.09**	**(0.03,0.15)**	0.04	(-0.02,0.10)
Match-N1	**-0.06**	**(-0.09,-0.02)**	-0.02	(-0.06,0.02)
Match-N2	**-0.05**	**(-0.09,-0.01)**	-0.02	(-0.06,0.02)
Amb*Match-N1	-0.06	(-0.14,0.02)	0.04	(-0.04,0.12)
Amb*Match-N2	-0.09	(-0.17,0.00)	-0.00	(-0.08,0.08)

Table 5.5: Bayesian linear mixed-effects models fit to log RTs at the critical and spillover regions of Experiment 7.

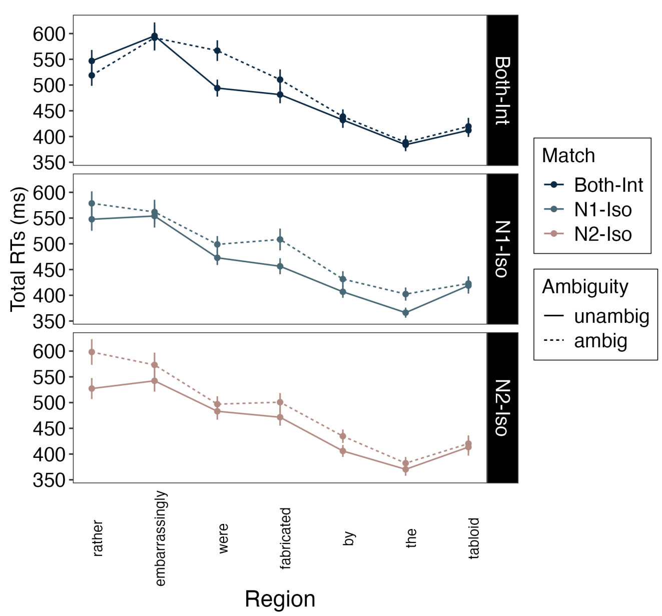
In first-pass times, RTs to the ambiguous BOTH-INT condition were longer than either of the ISO conditions at the critical region. Model results indicated main effects of Ambiguity ($\hat{\beta}$ = 0.05, CrI = [0.01, 0.1]) and Match (Match-N1: $\hat{\beta}$ = -0.05, CrI = [-0.08, -0.02]; Match-N2: $\hat{\beta}$ = -0.03, CrI = [-0.07, -0.00]), as well as a credible interaction between Ambiguity and Match (Amb*Match-N1: $\hat{\beta}$ = -0.06, CrI = [-0.12, -0.00]; Amb*Match-N2: $\hat{\beta}$ = -0.07, CrI = [-0.14, -0.01]). Notably, the Ambiguity x Match interactions suggest that the ambiguous BOTH-INT condition displayed a reanalysis effect at this region, but there was no numerical difference between RTs to the +AMBIG vs. -AMBIG ISO conditions. At the spillover region, there was a numerical trend such that



(a) First-pass times.



(b) Go-past times.



(c) Total times.

Figure 5.1: Word-by-word B-SPR latencies by condition for Experiment 7.

+AMBIG conditions were read slower than -AMBIG conditions, but this difference was not meaningful. There were no other credible differences at the spillover region.

Go-past times revealed a similar pattern of results. At the critical region, there were main effects of Match (Match-N1: $\hat{\beta}$ = -0.07, CrI = [-0.11, -0.03]; Match-N2: $\hat{\beta}$ = -0.05, CrI = [-0.1, -0.00]) such that the BOTH-INT condition was slower than either of the ISO conditions. There was a main effect of Ambiguity ($\hat{\beta}$ = 0.08, CrI = [0.02, 0.14]) such that RTs to +AMBIG > -AMBIG; once again, this was qualified by an interaction between Ambiguity and Match (Amb*Match-N1: $\hat{\beta}$ = -0.01, CrI = [-0.18, -0.02]; Amb*Match-N2: $\hat{\beta}$ = -0.12, CrI = [-0.21, -0.03]), where RTs to the Ambiguous INT condition were slower than those to the ambiguous ISO conditions. Here, too, the ISO conditions did not display evidence of a reanalysis cost, as the -AMBIG-ISO conditions were numerically slower than the +AMBIG ones. Again, there were no credible differences at the spillover region, although the +AMBIG RTs were numerically higher.

In total times at the critical region, model results revealed a main effect of Ambiguity ($\hat{\beta}$ = 0.09, CrI = [0.03, 0.15]) and main effects of Match (Match-N1: $\hat{\beta}$ = -0.06, CrI = [-0.09, -0.02]; Match-N2: $\hat{\beta}$ = -0.05, CrI = [-0.09, -0.01]), but no credible interaction. No credible effects emerged at the spillover region.

Curiously, the ISO conditions did not display a reanalysis cost at the critical region, and there was no credible main effect of Ambiguity or interaction between Ambiguity and Match at the spillover region, although +AMBIG-ISO structures displayed a numerical penalty relative to their -AMBIG-ISO baselines in this region. This led us to the concern that the ISO conditions may not have undergone reanalysis, especially in the N2-ISO case. The N1-ISO condition displayed a persistent numerical penalty for ambiguous constructions beyond the critical and spillover regions, suggesting the possibility that reanalysis proceeds slower in this condition, although attachment is facilitated, per the speed-up at the critical region. In contrast, the N2-ISO condition only displayed a non-credible ambiguity penalty at the spillover region in go-past and total times. In order to further investigate this observation, we conducted a post-hoc analysis on the two words after the spillover region (Spill2 = *by*, Spill3 = *the*). This analysis fit a `brms` model[47] to go-past times on these regions only for the ISO structures. Predictors were sum-coded, with the +AMBIG and N1-ISO conditions mapped to positive values. Results are in Table 5.6.

	Spill2 (*by*)		**Spill3** (*the*)	
Go-past RTs	$\hat{\beta}$	CrI	$\hat{\beta}$	CrI
Ambiguity	0.01	(-0.02,0.05)	0.002	(-0.04,0.04)
Match	-0.03	(-0.06,0.01)	0.03	(-0.01,0.07)
Amb*Match	0.01	(-0.08,0.1)	0.06	(-0.01,0.13)

Table 5.6: Post-hoc Bayesian linear mixed-effects models fit to log RTs at the Spill2 and Spill3 regions of Experiment 7.

Model results revealed no credible effects. We sought to assess the informativity of the Ambiguity*Match interaction at Spill3 in particular, where RTs revealed a crossover pattern such that N1-ISO displayed an ambiguity penalty, whereas N2-ISO showed facilitation for ambiguous structures. A Bayes Factor analysis compared interaction vs. non-interaction models using the `bridgesampling` package in R (Gronau et al., 2017), following the procedure specified by (Wagenmakers et al., 2018). This resulted in moderate evidence for the non-interaction model (BF_{10} = 0.13), based on Lee and Wagenmakers's (2014) classification scheme for interpreting Bayes Factors.

5.4.3 Discussion

The results are consistent with the predictions of Context-Sensitive Retrieval: RTs to both ambiguous ISO conditions were faster than to the ambiguous BOTH-INT condition,

[47] `brm(logRT ~ Ambiguity*Match + (1 + Ambiguity*Match | Subject) + (1 + Ambiguity*Match | Item))`

providing suggestive evidence for the claim that reanalysis proceeds faster in segmented structures. Additionally, the results suggest that accessing the second AWA coordinate in isolated structures is no more difficult than accessing the first AWA coordinate. This result provides further support for the claim that appositive relative clauses and AWA-coordination are alike in their processing profiles.

Moreover, the results did not provide support for Reinstantiation. Under this hypothesis, we predicted that reanalysis should be easier in the N1-ISO condition compared to the N2-ISO condition, because the content of the first coordinate is reinstated prior to the disambiguating input, which is a better match for N1 in the N1-ISO case. In other words, the feature-matching candidate in N1-ISO should lead to easier access to the first coordinate's encoding context. If anything, we saw a trend in the opposite direction: that reanalysis cost for N2-ISO was smaller than in N1-ISO, per the numerical increase in later regions in the ambiguous N1-ISO condition. It's also possible that Reinstantiation does not affect item-level retrieval in this way; that is, given the presence of informative features on the retrieval probe (in the absence of cue overload), the cue-based retrieval mechanism privileges those features, rather than relying on the contextual state. Therefore, N1-matching number features in the N1-ISO condition and N2-matching number features in the N2-ISO condition are privileged in guiding the search process, per Context-Sensitive Retrieval. We suggest, then, that the Reinstantiation process is still operative, but its effects are primarily observable in the presence of cue overload (when the retrieval probe provides a less informative cue).

Therefore, we take the results to be most consistent with CSR. There are two caveats to this conclusion, however. The first is that the results did not provide strong evidence that the ISO structures are adequately reanalyzed. This raises the possibility that attachment proceeds easily in these structures, per CSR, but reanalysis does not. This may be due to the fact that the presence of prosodic boundaries leads to stronger commitment to the incorrect parse, because prosodic phrases form interpretive domains that are difficult to revise once completed, per Schafer's (1997) suggestion. The current experiment cannot comment on this, because we did not collect substantial data on interpretations of the ambiguous region. Experiment 8 will address this issue.

Additionally, the lack of reanalysis effect in the ISO structures raises a question about the reading strategy comprehenders adopt in bidirectional SPR tasks. The incentive to reanalyze fully may be diminished in B-SPR, because readers know they can re-read earlier portions of the text. Of course, the ability to re-read is present in naturalistic reading contexts, but B-SPR is different from natural reading because readers must consciously choose to backtrack via button presses. Our data show a standard rate of regressions (on $\sim$10% of trials overall), but this does not mean that the accompanying processing strategy or the role of regressions in B-SPR is identical to that of eye-tracking while reading, for example (see Paape et al., 2022, for a more detailed discussion of B-SPR). All this aside, the BOTH-INT structures did display a clear reanalysis effect, so lack of a clear signature of reanalysis in the ISO structures likely cannot be attributed to the task alone.

The second note has to do with the baseline condition we used here. In the ISO structures, the auxiliary only matched one of the noun phrases, whereas in the INT structure, the auxiliary matched both. Therefore, the fact that we observed facilitation for ISO structures could be due to the fact that only the INT structures exhibited retrieval interference due to cue overload of plural features. We refute this explanation because our results showed main effects of Match in addition to the Ambiguity x Match interactions. Indeed, we observed an effect of retrieval interference within the -AMBIG conditions alone, such that RTs to the BOTH-INT condition were longer in the absence of reanalysis; these conditions are plotted separately in Figure 5.2. Additionally, visual inspection of these plots suggests a penalty for BOTH-INT at the pre-critical region (*embarrassingly*). It's possible that this reflects an early effect of retrieval interference for ordinary coordination. In conclusion, potential retrieval interference in the INT structure cannot entirely explain our results.

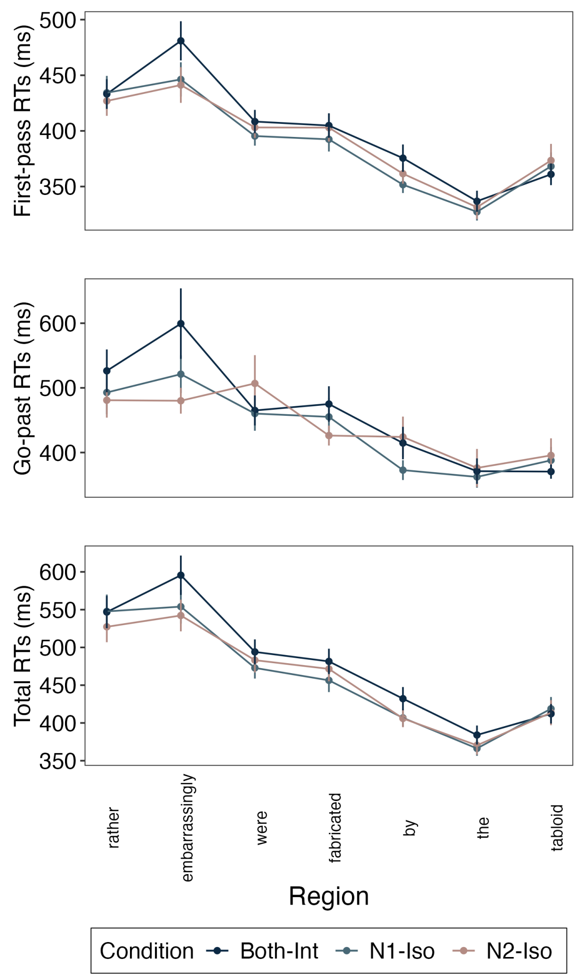


Figure 5.2: Word-by-word SPR latencies for -AMBIG conditions of Experiment 7.

Overall, we take the results to support Context-Sensitive Retrieval. But, in order to isolate the role of segmentation, a follow-up study could use the design in Table 5.7 to compare ease of reanalysis for coordinate one vs. coordinate two content in segmented and non-segmented structures in the absence of cue overload. In this design, CSR should predict greater ease of attachment and reanalysis for ISO structures compared to INT ones. I leave this possibility for future work.

Match	Structure	Ambiguity (∅, *that*) **critical** | spillover
MATCH-N1	ISO	Ramona believed {∅, that} the rumors, as well as the article, **were** fabricated by the tabloid.
	INT	Ramona believed {∅, that} the rumors and the article **were** fabricated by the tabloid.
MATCH-N2	ISO	Ramona believed {∅, that} the rumor, as well as the articles, **were** fabricated by the tabloid.
	INT	Ramona believed {∅, that} the rumor and the articles **were** fabricated by the tabloid.

Table 5.7: Example item set for proposed follow-up to Experiment 7.

.

5.5 Experiment 8

Experiment 8 aimed to replicate the reading time results of Experiment 7 while probing rates of semantic persistence for the contents of each coordinate. Recall that Context-Sensitive Retrieval should predict a lesser rate of semantic persistence for the first coordinate in ambiguous N1-ISO structures and lesser persistence for the second coordinate in ambiguous N2-ISO structures. In contrast, the Interpretive Domains Hypothesis (Schafer, 1997) should predict equally difficult reanalysis for both ISO structures (i.e., a main effect of Ambiguity, and no interaction with Match), and perhaps greater difficult reanalyzing N1-ISO, if its incorrect analysis is reinforced, per Reinstantiation. A pure effect of Reinstantiation should predict easier reanalysis for ambiguous N1-ISO structures only, if access to the reinstated context is facilitated. These predictions are outlined in Table 5.3.

5.5.1 Method

5.5.1.1 Participants

110 participants were recruited from the UC Santa Cruz Linguistics Department Subject Pool to participate in the experiment. All participants self-reported that they began learning English at or before the age of 7. Each experiment session took approximately 60 minutes, and participants were compensated with course credit. 9 participants were excluded due to average comprehension question accuracy below 70%. The analysis included data from 101 participants.

5.5.1.2 Materials

64 item sets crossing Ambiguity (+/-AMBIG), Match (N1, N2), and Question-Target (N1-Q, N2-Q) were constructed. The current experiment used only isolated AWA-coordination structures. Materials were identical in nature to the ones used in Experiment 7: they contained AWA-coordination with NP/S ambiguities (+/-AMBIG), where the disambiguating verb either matched N1 or N2. In addition, the Q-Target factor varied whether an end-of-sentence comprehension question targeted the content of the first (N1-Q) or second (N2-Q) coordinate. Comprehension questions always asked about the ambiguous region in order to probe the degree of semantic persistence in each condition. Therefore, the question always asked whether the incorrect analysis was true (*Did Ramona believe N1/N2?*), to which the correct answer was always "no". A sample item set is provided in Table 5.8.

The ambiguous verbs used were the same as in Experiment 7. The current experiment also attempted to ensure that the NP-complement parse was rendered implausible at the embedded verb region, following Sturt (2007). In the current experiment, we removed the pre-critical adverbial region present in Experiment 7, because the results

Match	Ambiguity (+AMBIG = ∅, -AMBIG = *that*)	**critical** | spillover
MATCH-N1	Ramona believed {∅, that} the rumors, as well as the article, **were** fabricated by the tabloid.	
Q-Target	N1-Q: Did Ramona believe the rumors?	
	N2-Q: Did Ramona believe the article?	
MATCH-N2	Ramona believed {∅, that} the rumor, as well as the articles, **were** fabricated by the tabloid.	
Q-Target	N1-Q: Did Ramona believe the rumor?	
	N2-Q: Did Ramona believe the articles?	

Table 5.8: Example item set from Experiment 8.

.

of that experiment suggested that this region may have encouraged participants to engage in subject retrieval prior to encountering the verb.

5.5.1.3 Procedure

The procedure was identical to that of Experiment 7, except that participants read sentences using a standard moving window self-paced reading paradigm, not a bidirectional one. This change to the procedure was made because of the concern about B-SPR changing participants' reading strategies, and because we sought to determine whether the initial stages of cue-based retrieval could influence readers' interpretations. Therefore, we attempted to eliminate the ability to re-read sentences to facilitate reanalysis. In addition, participants responded to a comprehension question following every sentence in the current experiment.

5.5.2 Results

Average comprehension question accuracy was 74% (56% for the experimental items and 91% for the fillers). Accuracy on the experimental items was surprisingly low. This is likely due to the fact that all experimental items prosodically isolated the temporarily ambiguous region, and all comprehension questions for the experimental items targeted this region. Compared to Experiment 7, comprehension question accuracy on the fillers was quite high, although they included ungrammatical sentences with agreement errors and NP/Z garden path constructions, like in the previous experiment. This suggests to us that the experimental items were particularly difficult for participants, or some participants displayed a "yes" bias, as the correct answer to the filler comprehension questions was always "yes", and the correct answer to the experimental comprehension questions was always "no". We discuss this further in the context of the results. Self-paced reading results are in §5.5.2.1, and analysis of semantic persistence rates is in §5.5.2.2.

5.5.2.1 Self-Paced Reading Results

Word-by-word RTs to each condition are plotted in Figure 5.3.

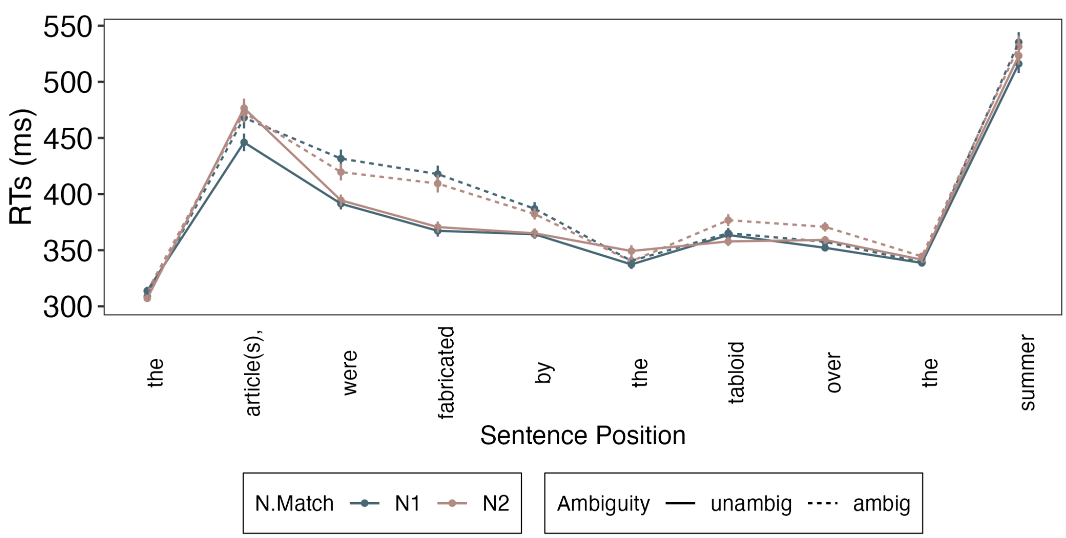


Figure 5.3: Word-by-word SPR latencies by condition for Experiment 8.

We fit Bayesian linear mixed-effects models[48] using the brms package in R (Bürkner,

[48]`brm(logRT ~ Ambiguity*Match + (1 + Ambiguity*Match | Subject) + (1 + Ambiguity*Match | Item))`

2017) to the critical and spillover regions. Predictors were sum-coded, with -AMBIG and N2 conditions mapped to negative values. $\widehat{R}$ values and posterior predictive checks indicated model convergence (Gelman et al., 2014). Model results are in Table 5.9.

	Critical (*were*)		Spillover (*fabricated*)	
Effect	$\widehat{\beta}$	CrI	$\widehat{\beta}$	CrI
Ambiguity	**0.04**	**(0.01,0.07)**	**0.07**	**(0.04,0.09)**
Match	0.001	(-0.02,0.02)	0.005	(-0.02,0.02)
Amb*Match	0.02	(-0.03,0.07)	0.03	(-0.02,0.08)

Table 5.9: Bayesian linear mixed-effects models fit to log RTs at the critical and spillover regions of Experiment 8.

At the critical region, +AMBIG conditions displayed a penalty relative to -AMBIG conditions. This resulted in a main effect of Ambiguity ($\widehat{\beta}$ = 0.04, CrI = [0.01, 0.07]). In addition, the ambiguous MATCH-N1 condition displayed a numerical penalty relative to ambiguous MATCH-N2 condition. However, the interaction between Ambiguity and Match was not borne out in the model results ($\widehat{\beta}$ = 0.02, CrI = [-0.03, 0.07]).

A main effect of Ambiguity also emerged at the spillover region ($\widehat{\beta}$ = 0.07, CrI = [0.04, 0.09]). Once again, this was accompanied by a numerical trend such that ambiguous MATCH-N1 RTs were slower than ambiguous MATCH-N2 RTs, but this interaction was not credible ($\widehat{\beta}$ = 0.03, CrI = [-0.02, 0.08]).

The ambiguous MATCH-N2 condition displayed inflated RTs 3-4 words past the spillover region (Spill4 = *tabloid*, Spill5 = *over*). This trend, along with the numerical speed-up for MATCH-N2 at the critical and spillover regions, potentially indicated easier attachment for this condition, but a larger reanalysis cost further downstream. To investigate this pattern further, we conducted a post-hoc analysis on these regions. We fit `brms` models to both regions, with the same specifications as those fit to the critical and spillover regions. These results are in Table 5.10.

Only a main effect of Ambiguity emerged at the Spill4 ($\widehat{\beta}$ = 0.03, CrI = [0.006, 0.05])

Effect	Spill4 (*tabloid*) $\hat{\beta}$	Spill4 (*tabloid*) CrI	Spill5 (*over*) $\hat{\beta}$	Spill5 (*over*) CrI
Ambiguity	**0.03**	**(0.006,0.05)**	**0.01**	**(0.0007,0.03)**
Match	-0.007	(-0.02,0.01)	-0.02	(-0.03,0.002)
Amb*Match	-0.01	(-0.05,0.03)	0.002	(-0.03,0.04)

Table 5.10: Post-hoc Bayesian linear mixed-effects models fit to log RTs at the Spill4 and Spill5 regions of Experiment 8.

and Spill5 ($\hat{\beta}$ = 0.01, CrI = [0.0007, 0.03]) regions. The interaction between Ambiguity and Match was not meaningful in either region.

5.5.2.2 Comprehension Question Results

Recall that Context-Sensitive Retrieval hypothesis predicts a 3-way interaction between Ambiguity, Match, and Q-Target such that the proportion of "no" (correct) responses should be higher for the first coordinate (N1-Q) in the ambiguous MATCH-N1 condition and the proportion of "no" responses should be higher for the second coordinate (N2-Q) in the ambiguous MATCH-N2 condition.

The proportion of "no" responses by condition is plotted in Figure 5.4. We fit a Bayesian logistic mixed-effects model[49] to the data. Predictors were sum-coded, with +AMBIG, MATCH-N1, and N1-Q conditions mapped to positive values. Model results are in Table 5.11.

Results revealed a credible effect of Ambiguity ($\hat{\beta}$ = 0.62, CrI = [0.64, 1.0]) such that accuracy for +AMBIG conditions was worse than -AMBIG conditions. In general, accuracy for +AMBIG conditions was close to chance, suggesting that participants experienced significant difficulty reanalyzing these structures. No other meaningful effects emerged, but I note some trends in the data. Responses to N1-Q questions were generally less accurate than responses to N2-Q questions ($\hat{\beta}$ = -0.16, CrI = [-0.33, 0.002]).

[49] `brm(Accuracy ~ Ambiguity*Match*Q-Target + (1 + Ambiguity*Match*Q-Target | Subject) + (1 + Ambiguity*Match*Q-Target | Item))`

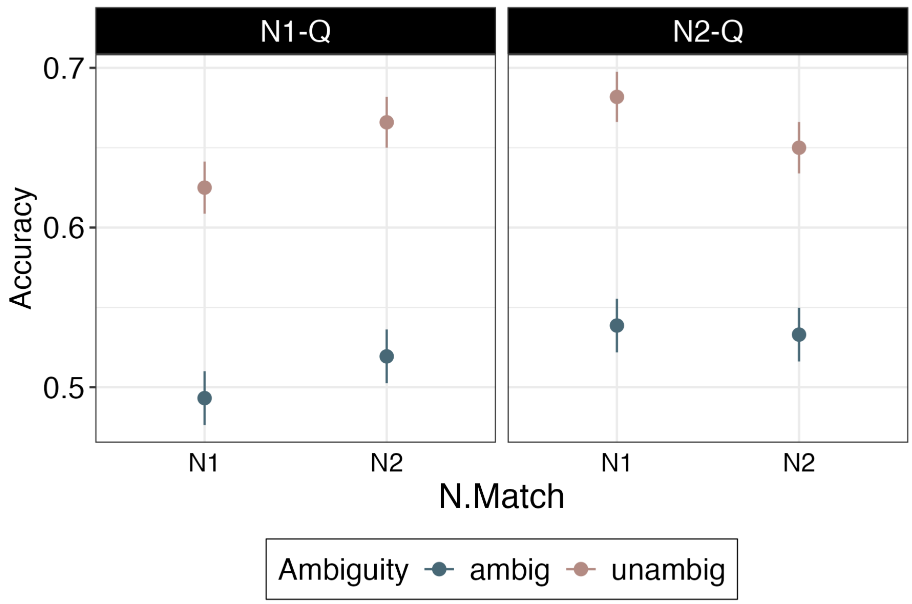


Figure 5.4: Comprehension question accuracy by condition for Experiment 8, indicating the proportion of "no" responses.

In addition, the interaction between Match and Q-Target suggested a trend such that accuracy for MATCH-N1 was lower than MATCH-N2 in the N1-Q conditions, whereas accuracy for MATCH-N1 was higher than MATCH-N2 in the N2-Q conditions ($\hat{\beta}$ = -0.16, CrI = [-0.33, 0.002]). Relatedly, mean accuracy for the ambiguous MATCH-N1 structure within the N1-Q condition was below chance. This suggests that this structure resulted in worse accuracy specifically when the comprehension question targeted the first coordinate. I discuss this trend further in the discussion.

Although the 3-way interaction was not credible, the results suggest that reanalysis of the second coordinate was more successful given a matching N2; this is predicted under CSR. On the other hand, reanalysis of the first coordinate was unsuccessful given a matching N1, as evidenced by the at-chance N1-Q responses. In order to assess evidence for the critical interaction predicted under CSR, we conducted a Bayes Factor analysis comparing a model containing the 3-way interaction between Ambiguity, Match, and Q-Target to a reduced model without this interaction, per the procedure in

Effect	$\hat{\beta}$	CrI
Ambiguity	**0.62**	**(0.64,1.0)**
Match	-0.6	(-0.2,0.07)
Q-Target	-0.16	(-0.33,0.002)
Amb*Match	-0.02	(-0.33,0.29)
Amb*Q-Targ	0.07	(-0.19,0.33)
Match*Q-Targ	-0.26	(-0.59,0.07)
Amb*Match*Q-Targ	-0.29	(-0.83,0.25)

Table 5.11: Bayesian logistic mixed-effects model on proportion "no" responses to comprehension questions from Experiment 8.

Wagenmakers et al. (2018). This analysis suggested moderate evidence in favor of the 3-way interaction model (BF_{10} = 3.1), per Lee and Wagenmakers's (2014) classification scheme. We discuss this trend further in the discussion.

Due to the fact that we observed low accuracy in the +AMBIG conditions overall, we split the results by lower- vs. higher-accuracy subjects (those whose accuracy was below vs. above the overall median accuracy (81%), respectively), because we hypothesized that these groups may have engaged in different reanalysis strategies, if success of reanalysis is dependent on comprehenders' ability to make use of available retrieval cues. Results are plotted in Figure 5.5. There are too few observations to statistically assess the 3-way interaction within each subset, but within the high-accuracy group, the data trends more clearly in the direction discussed above: that reanalysis of the second coordinate was more successful given a matching N2, and in fact, reanalysis of the first coordinate is slightly less successful given a matching N1. Within the low-accuracy group, it seems that reanalysis of the first coordinate was more successful given a matching N1, whereas reanalysis of the second coordinate was slightly less successful given a matching N2. However, the low-accuracy group is so close to chance that it is difficult to interpret this pattern; we suspect that performance on the +AMBIG conditions here may reflect a floor effect.

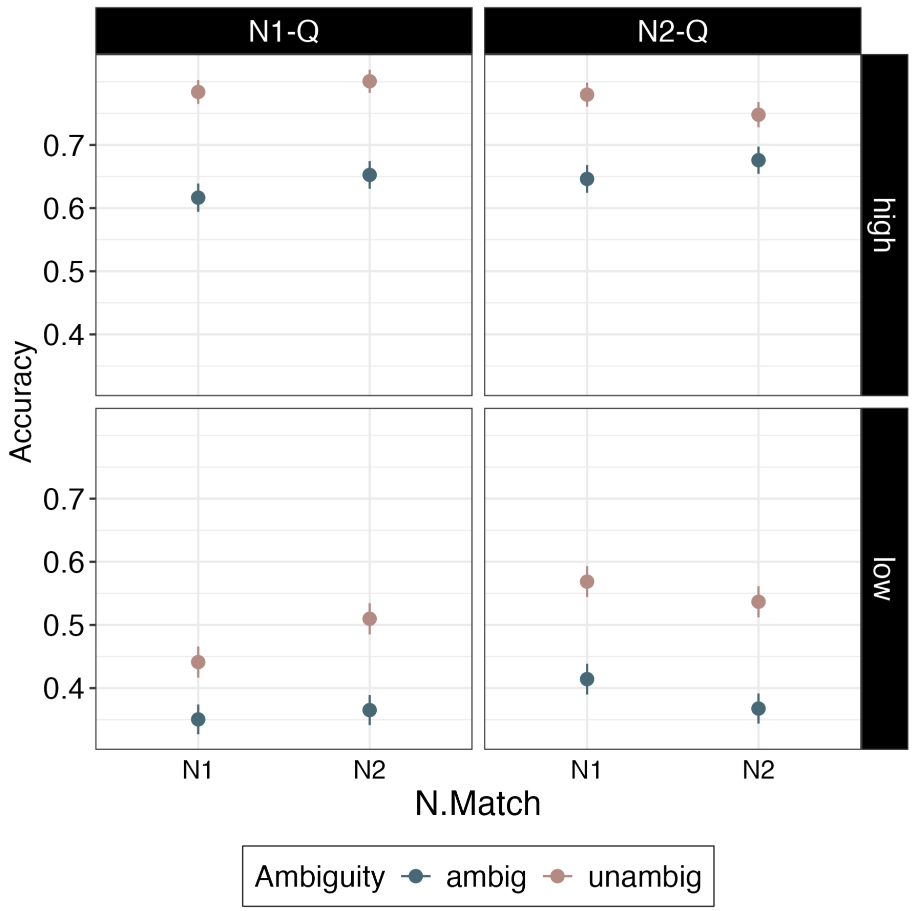


Figure 5.5: Comprehension question accuracy by condition by accuracy level for Experiment 8, indicating the proportion of ”no” responses.

5.5.3 Discussion

The SPR results revealed a main effect of Ambiguity, suggesting good evidence that reanalysis was attempted for isolated AWA-coordination in the current experiment, unlike in Experiment 7. In addition, there was a trend towards the reanalysis cost for the MATCH-N2 condition being slightly smaller at the critical and spillover regions. Although this did not result in a credible interaction between Ambiguity and Match, this trend was accompanied by inflated RTs to the ambiguous MATCH-N2 condition later on. This pattern hints at easier attachment within an AWA-coordinate-internal site, but greater difficult with reanalysis. However, the model results did not validate this pattern in any region.

The current experiment primarily sought to investigate the rate of semantic persistence for each condition, in order to determine whether retrieval of a subject in segmented structures requiring reanalysis facilitated access to the retrieved item's encoding context. Under CSR, we predicted that this would have resulted in greater success of reanalysis for the targeted coordinate, in contrast to the non-targeted one. The model results did not provide strong evidence for any of the hypotheses we entertained, as we obtained only a main effect of Ambiguity. But, the Bayes Factor analysis suggested moderate evidence in favor of an interaction, so we consider some speculative explanations here.

In particular, for responses to comprehension questions targeting the second coordinate, the results were suggestive of CSR, because reanalysis cost for the second coordinate was reduced when the retrieval probe matched N2 in number features. This pattern was especially evident in the high accuracy group (Figure 5.5). For comprehension questions targeting the first coordinate, however, the predictions of CSR were not borne out. Within the N1-Q conditions, only a main effect of Ambiguity (+AMBIG > -AMBIG) and a penalty for MATCH-N1 conditions emerged. Within the high-accuracy group, there was a numerical trend such that accuracy for the ambiguous MATCH-N1 condition was worst. This pattern is more consistent with the predictions of Reinstantiation + Interpretive Domains, which assumes that prosodic boundaries encourage deeper commitment to the incorrect analysis in ambiguous conditions, and that this commitment may be further reinforced by reinstating the encoding context of the first coordinate. The results then tentatively suggest a pattern more consistent with Reinstantiation/Interpretive Domains for N1-Q conditions, but a pattern more consistent with CSR for N2-Q conditions. This may be because at the time of encountering the critical region, readers have encountered two preceding prosodic boundaries that may serve to strengthen the incorrect representation of the first coordinate. Furthermore, reinstating the initial encoding context prior to the disambiguating region may strengthen the representation of that context in memory, especially if a wrap-up process where comprehenders integrate that context's content into the discourse representation takes place. In addition, when asked *Did Ramona believe the rumors?*, surface match between the question and their gist representation of the first coordinate may make participants more likely to incorrectly respond "yes". This accords with the view that reanalysis is particularly difficult when it spans one or more prosodic boundaries. It also explains why participants were more likely to incorrectly respond to a MATCH-N1 condition when asked about the first coordinate than when asked about the second coordinate. In response to N2-Q, however, perhaps less time passed and a lesser degree of commitment to the incorrect parse allows the parser to more easily utilize retrieval cues available on the disambiguating input in order to correct the misanalysis. We suggest then that the results provide suggestive support for both the Interpretive Domains Hypothesis and CSR, but future work should aim to understand how comprehenders mediate between different cues during syntactic analysis.

It is noteworthy that the accuracy for ambiguous constructions in the current experiment was very low. It's possible that the difficulty of the structures combined with the fact that the comprehension questions were asked long after the disambiguating input was encountered resulted in lack of valid cues for participants to adequately engage in reanalysis. In this way, our results mirror other work that suggests separable effects of attachment and reanalysis. While attachment may proceed smoothly, and was even facilitated in our segmented structures, revision of the initial parse was more difficult. One caveat to this conclusion is that our comprehension questions did not probe whether participants successfully arrived at the correct interpretation. However, we assume based on the pattern of SPR results that participants engaged in revisionary processes at the critical and spillover regions.

5.6 General Discussion

We set out to test the predictions of CSR for a linguistic dependency for which retrieved context is more likely to matter, namely syntactic reanalysis. For sentences requiring reanalysis, we conjectured that retrieval of a feature-matching candidate in isolated AWA-coordination would facilitate access to the syntactic structure local to the retrieved item, but inhibit access to the structure of the other coordinate. We further predicted that better access to a retrieved item's encoding context would lead to greater success with reanalysis. The B-SPR results of Experiment 7 clearly supported the role of CSR: we found that ISO structures yielded a smaller difference between +/-AMBIG conditions than INT ones. However, the interpretive data reported in Experiment 8 did not support our predictions about reanalysis. In fact, the low accuracy for +AMBIG conditions suggested that comprehenders had substantial difficulty reanalyzing the structures used here. Taken together, these results suggest that segmented structures facilitate attachment of disambiguating input, but not revision of previously misanalyzed content.

The results lead to an interesting conclusion: that prosodic boundaries play a dual role in (i) partitioning syntactic content in memory into encoding contexts and (ii) delineating interpretive domains, perhaps at a higher level of linguistic representation. Perhaps unsurprisingly, it seems that cue-based retrieval operations are able to make use of the former domains (encoding contexts) but not of the latter (discourse segments) in order to facilitate linguistic processing. Perhaps the issue of whether the parser opts to access a level of representation that is sensitive to encoding contexts is dependent on the particular dependency at hand. For example, syntactic dependencies like agreement and attachment seem to be sensitive to encoding context boundaries, per Context-Sensitive Retrieval. In contrast, ellipsis did not display a clear effect of CSR in Chapters 3-4. It's notable that the effect of Reinstantiation appears to emerge for syntactic (Experiment 5) and discourse-level dependencies (S. Kim & Xiang, 2023) alike. Perhaps, then, Reinstantiation as a mechanism is more robust to dependency type, whereas CSR primarily emerges for syntactic dependencies that require access to larger pieces of syntactic structure.

This difference is captured under the proposal put forward in §4.4. If Reinstantiation is a process driven by the need to reactivate an earlier context so that it can be used for an upcoming subject-verb dependency, this should occur very reliably in structures where the subject is prosodically separated across multiple segments. In turn, it will also make the reinstated context more accessible for any subsequent dependency (agreement, filler-gap integration, anaphora, etc.). However, a retrieval cue that better matches content in an earlier encoding context should always be utilizable. This explains the fact that the content internal to ARCs and the second coordinate of AWA is equally accessible compared to non-segmented baseline structures, which refutes the predictions of Visibility. Overall, then, we do not find evidence supporting Visibility for item-to-item dependency resolution (if we further attribute the results of Experiment 4 to the alternative focus-based explanation we put forward; see §4.1.4). This further suggests that where Visibility effects do arise, these are not due to the organization of prosodic phrases in memory, contra CSE-Visibility. On the other hand, Experiment 7 supports a limited role for Context-Sensitive Retrieval, only for processes that require accessing previously built syntactic structure.

Chapter 6

Conclusion

At a high level, the conclusions of the dissertation are straightforward: linguistic segmentation has a measurable effect on online sentence comprehension, both at boundary positions and during later processing. This suggests that the role of segmentation in lingusitic memory is not limited to sentence-final representations. Instead, it would appear that certain segments fluctuate in activation throughout the timecourse of processing a sentence, and that the cue-based retrieval mechanism is sensitive to these fluctuations. I've further argued that these fluctuations are not random, but rather are driven by pre-activation of particular lingusitic content, like (part of) a subject, in advance of an upcoming dependency. Additionally, the results suggest that at least for some dependencies, the presence of prosodic segments serves to reduce interference (from non-targeted encoding contexts) during cue-based retrieval, suggesting that a particular segment can sometimes be accessed to the exclusion of other sentence content. In this way, the Context-Sensitive Encoding hypothesis advanced here provides a means of (i) navigating the global linguistic structure and (ii) mediating access to different segments during online sentence comprehension.

The dissertation established these facts using three types of prosodically segmented constructions as case studies: appositive relative clauses, and two types of focus-sensitive coordinate structures (*not only...but also* and *...as well as...*). In each of these cases, syntactic, prosodic, and discourse boundary positions align. However, the results established that prosodic boundaries in particular play a special role in partitioning syntactic content into encoding contexts in memory. Across the 8 experiments, the results pointed to slightly different conclusions about the mechanisms that mediate the relationship between boundaries and the particular dependency at hand. Consequently, I summarize the findings of each chapter below in an attempt to clarify some of these differences.

6.1 Overview of Findings

The main findings are summarized in Table 6.1.

❧ **Chapter 2** reviewed literature on the effect of boundaries in linguistic memory, and in the process, revealed certain inconsistencies in the assumptions that different psycholinguistic studies have adopted about the nature of memory at different levels of representation. It then motivated the idea that a notion of temporal context can allow the parser to mark structural boundaries during encoding. In this way, the contextual features of a particular segment can be reinstated at certain points, and interference from other segments can be reduced.

❧ **Chapter 3** revealed that while appositive relative clauses may sometimes be backgrounded in the discourse, their representation persists in memory. Backgrounding and Compression-based accounts of ARCs have attributed their tendency to be bypassed online to their backgrounded discourse status, and thus predict (i) worse memory for ARC content and (ii) worse access to ARC-internal content. Experiments 1-3 refuted these predictions. Experiment 1 evidenced better retention of syntactic memory for ARCs compared to restrictive relative clauses, in line with the Context-

Sensitive Encoding hypothesis. Together, Experiments 2-3 suggested no difference in access to ARC-internal vs. RRC-internal content. On conceptual grounds, the experiments suggested that Reinstantiation offers a better account of ARC bypassing.

❧ **Chapter 4** empirically validated the role of Reinstantiation for medial discourse segments that are not backgrounded, in *as well as* coordinate structures. In other words, we obtained evidence of bypassing for both ARCs and AWA-coordination, and therefore surmised that bypassing of ARCs is not due to their differential discourse status. We further found that spending more time on a prosodically marked boundary relates to a benefit in memory for the content of the preceding segment as a whole. This further substantiated Context-Sensitive Encoding. In the investigation of *not only...but also* structures, we found evidence more in line with Visibility than Context-Sensitive Retrieval (retrieval across a medial prosodic boundary incurred a penalty), but we ultimately attributed this to differences in focus structural expectations for the two constructions we contrasted (NOBA vs. *and also*), driven by the fact that NOBA contains an initial focus-sensitive coordinator. Overall, then, the results did not provide evidence for Context-Sensitive Retrieval or Visibility in the resolution of item-to-item dependencies.

❧ **Chapter 5** found evidence for Context-Sensitive Retrieval in a particular context: that of syntactic reanalysis in isolated AWA coordinate structures. We argued that when the dependency cueing retrieval requires accessing previously built syntactic structure from memory, not just word-level features, the amount of contextual interference within the target encoding context does have an effect. Furthermore, our results suggested that while this process evidences facilitation for segmented structures during reading, this did not relate to successful comprehension. This relates to the finding that the effects of attachment and reanalysis are dissociable (Van Dyke & Lewis, 2003): while prosodically segmented structures may facilitate attachment processes, they do not appear to facilitate reanalysis (in fact, they may have even hindered it). Under Context-Sensitive Retrieval, we predicted that a better match between retrieval cues on the disambiguating region and the features of a particular coordinate may benefit reanalysis of that coordinate alone. Insofar as we found any evidence for that, it was only suggestive, and only within coordinate two. We concluded that interpretive domains (c.f. Schafer, 1997) formed by prosodic boundaries inhibit the parser's ability to effectively utilize item-level cues to guide reanalysis.

6.2 Future Directions

The current work presents a number of directions for future investigations. I briefly entertain some possibilities here.

❧ **Replications in listening**

The current work has adopted the Implicit Prosody Hypothesis, and has argued for the role of implicit prosodic boundaries in delineating encoding contexts. In order to validate the claim that the effects reported here are due to prosodic boundary positions in particular, these experiments should be replicated as listening studies. A significant challenge with extending this work to listening, particularly for ARCs, is understanding how the intonation of appositives affects their discourse status. At present, we know very little about how different appositive intonations may relate to the discourse relations comprehenders posit, but Auran and Loock (2011) suggest that certain types of appositives (namely continuative ones) feature fewer correlates of prosodic separation than others (e.g., Subjectivity and Relevance ARCs). Similarly, the intonational contour assigned to *as well as* coordinates may determine whether their content is interpreted as a discourse aside. To my knowledge, no previous work addresses the discourse-prosody relationship for AWA coordination. Listening stud-

Struc.	Summary	CSE	Rein.	CSR	Comp.	Back.	Vis.	IDH
ARC	**E1**: ARC content is better remembered than RRC content.	✔	–	–	✗	–	–	–
	E2: Access to main clause content is facilitated in sentences with final ARCs.	✔	–	?	✗	✗	✗	–
	E3: ARC-internal content is no less accessible than RRC-internal content, and ARC structures show late facilitation.	✔	–	✗	✗	✗	–	–
NOBA	**E4**: Inhibition (?) of dependencies spanning sentence-medial prosodic boundaries.	–	–	✗	–	–	?	–
AWA	**E5**: Isolated AWA structures evidence bypassing.	✔	✔	–	–	✗	✗	–
	E6: Integrated AWA does not evidence bypassing, and longer wrap-up at boundaries improves memory.	✔	✔	–	✗	–	✗	–
	E7: Access to isolated AWA content (coordinates one and two) is facilitated.	✔	–	✔	✗	✗	✗	–
	E8: Tentatively, coordinate two of isolated AWA is better reanalyzed than coordinate one.	–	–	?	–	–	–	✔

Table 6.1: A summary of evidence for proposed mechanisms across the 8 experiments. ***Key***: ARC = isolated appositive relative clauses, compared to integrated restrictive relative clause baselines; NOBA = separated *not only...but also* compared to integrated *and also* baselines; AWA = isolated *as well as* compared to integrated *and* or integrated AWA baselines.

ies should therefore be preceded by a substantial amount of linguistic groundwork to clarify these issues.

❧ **Cross-linguistic extensions**

In relation to the previous point, future work should focus on the processing profile of prosodically segmented constructions in other languages, in both reading and listening. If the prosody-memory relationship is central to forming encoding contexts in some way (c.f. Frazier et al., 2006), we should see similar segmentation and bypassing effects emerge cross-linguistically. However, we should also expect language-specific prosodic properties to guide the formation of encoding contexts. Few cross-linguistic investigations of the prosody of appositives exist. A comparison of German and Dutch suggests that Dutch listeners utilize temporal and intonational cues to distinguish between appositive and restrictive relative clauses, whereas German listeners do not (Kaland & Heuven, 2010). The study relates this to the orthographic convention in German to mark both appositive and restrictive relative clauses with commas. Investigating the processing of appositives in Dutch and German may shed light on the extent to which the bypassing effect is dependent on comma-marked boundaries.

In addition, the syntax of appositives has been argued to differ significantly across languages (Cinque, 2006; Cinque, 2020; De Vries, 2023). Some work claims that in languages like Mandarin and Japanese, for example, appositive relative clauses are syntactically integrated. Moreover, it is argued that some languages (Italian, Catalan) have both unintegrated and integrated appositives. If unintegrated syntactic units correlate with prosodic separation as well – like in the case of other constructions (c.f. Frey & Truckenbrodt, 2015) – it may be that languages like Italian have both prosodically integrated and isolated appositives. If a prosodic division between types of appositives holds in some languages, the current account then makes a testable prediction, as I've suggested that the initial partitioning of ARCs in memory at the time of encoding is influenced by their prosodic isolation. This would predict that isolated but not integrated ARCs in languages like Italian should evidence bypassing.

❧ Computational models of the TCM

The current work has laid out a conceptual sketch of the application of the Temporal Context Model to sentence memory. However, the TCM is a mathematical framework that has several concrete computational implementations (Howard & Kahana, 2002; Polyn et al., 2009; Pu, Kong, Ranganath, & Melloni, 2022b; Sederberg et al., 2008). Such models can offer an explicit means of testing the predictions of the account put forward here. In particular, one study (Pu et al., 2022a) developed a computational implementation of TCM to capture context "reset" at event boundaries, based on data from a series of experiments on temporal order memory within and across events showing that within-group memory was enhanced. Specifically, their model assumed that at non-boundary timepoints, the contextual representation gradually fluctuates in a manner analogous to Howard and Kahana's (2002) contextual evolution equation (see §2.3.2); then at each boundary position, some proportion of the pre-experimental context is reinstated. A similar framework could be applied to capture the bypassing effect. One caveat is that this cognitive modeling should be done in conjunction with additional experimentation, in order to better understand the relationship between temporal context and sentence memory.

❧ Syntactic domains and temporal context

Chapter 4 briefly discussed whether forward-looking syntactic dependencies affect context encodings, a point that requires elucidation in the current account. Note that appositive relative clauses always involve a forward-looking search process that spans their left boundary, because there is a syntactic relationship between the RC head and the RC-internal gap site. But, this does not appear to drive contextual likeness between the initial segment of the main clause and the ARC. Because ARCs also feature prosodic boundaries which drive isolation between encoding contexts, it would be useful to investigate the contrast between active search vs. retrieval using non-segmented structures. This would allow us to uncover additional information about how syntactic domains may serve to delineate encoding contexts (see Wagers, 2008).

The introduction alluded to other work on syntactic processing that shows that retrieval is sensitive to structural domains. Notably, comprehenders are very unlikely to erroneously retrieve structurally illicit distractors while searching for the antecedent of a reflexive pronoun (Dillon et al., 2013), whereas such errors abound for other types of linguistic dependencies, like subject-verb agreement (Wagers et al., 2009).

In (1), the NP inside the relative clause is not accessible to bind the pronoun *themselves* in the main clause, because its C-COMMAND domain[50] (Chomsky, 1981; Reinhart, 1983) does not contain the pronoun. Although there is featural overlap between the illicit NP (*the middle managers*) and the reflexive (i.e., the +PLURAL feature), the presence of this distractor does not ameliorate the penalty at the reflexive in a sentence like (1) during reading. This has led to the suggestion that structural cues to an antecedent outweigh morphological cues in these particular cases, but a significant challenge has been determining how to mechanistically account for the role of structure, because relational properties like C-COMMAND cannot be captured via item encodings (see Kush, 2013, for an extensive discussion).

(1) *The new executive $\boxed{\text{who oversaw the middle managers}}$ apparently doubted **themselves** on most major decisions.

Subsequent cross-linguistic studies have come to a slightly different conclusion: that in reflexive processing in Mandarin, for example, there is generally a preference

[50] A category A c-commands a category B if A does not dominate B and the first node dominating A also dominates B.

to search the most local syntactic domain for an antecedent (Dillon et al., 2014). This may be obscured in English, because the verb reactivates the licit antecedent immediately prior to encountering the pronoun (Pizarro-Guevara & Dillon, 2022). Indeed, English is more prone to show interference from a distractor when the reflexive is distanced from the verb (King, Andrews, & Wagers, 2012), and interference from the local domain emerges even in languages like Telugu, where there is a grammatical requirement for an anti-local antecedent (Arvindam & Wagers, 2022).

Recall that Wagers (2008) originally proposed the TCM could be used to mark clausal boundaries. Under this view, a notion of temporal context could also be used to explain the profile of reflexive processing. If a reflexive prefers to search for an antecedent in the most local syntactic domain, this could relate to a preference to access the most recent encoding context that contains a partially feature-matching, potential candidate. Such an approach could avoid the issue of having to posit a structural feature on each item, but still capture Dillon et al.'s proposal that the search for an antecedent is constrained by positional syntactic cues.

❧ Metacognition at linguistic boundaries

The current work sought to investigate the relationship between wrap-up and memory, but it also alluded to a connection between boundaries in text and certain nonlinguistic processes, like self-regulating comprehension. I suggested in §4.3.1 that this would be another fruitful area for future work. Generally speaking, if psycholinguists are to reason about the relationship between processing time and linguistic operations in a principled way, we should also seriously consider factors related to task demands, decision making, and metacognitive processing. These factors each relate to rich literatures of their own, and are bound to interact with linguistic memory. If, for example, there are individual differences in whether readers opt to use dwell time at boundaries to "consolidate" segments in memory, integrate segments into a discourse representation, or self-reflect on comprehension, this may also lead to differential reanalysis profiles. Those readers who allocate more effort to "self-checking" (Weiss et al., 2018) may be better able to recover from misanalyses than readers who expend effort on integrating content into the discourse, if Schafer's (1997) proposal that certain boundaries encourage deeper conceptual processing is true. The current work has left the nature of these interactions underspecified, but my eventual hope is that future investigations can provide us with a more holistic outlook on comprehending language, one that includes an understanding of both linguistic and metacognitive processes.

6.3 Final Wrap-Up

Broadly, I've motivated here that the presence of higher-order linguistic structure matters in offline sentence memory and during dependency resolution in sentence processing. This conclusion will likely come as no surprise to linguists, whose primary concern is the hierarchical nature of structure in language. But, I hope to have further motivated that psycholinguists should continue looking beyond cue-based retrieval – this model alone cannot universally account for linguistic processes. Rethinking the mechanism by which linguistic structure interacts with memory served to illuminate the workings of a particular psycholinguistic phenomenon in the current work, but at a higher level, it highlights the way in which parsing and search operations might depend on one another, although investigations of each area are often kept separate.

Previous work reviewed in this dissertation also exposes a somewhat pervasive claim throughout the psycholinguistic literature: that at certain points during the processing of a sentence, memory for previous content is lost (Carlson et al., 2009; Dillon et al., 2017; Frazier et al., 2014; Parker & Phillips, 2016). This claim likely has roots in the longstanding idea that memory for surface detail rapidly degrades, leaving some more durable but less granular gist representation in its place (Potter & Lombardi, 1990; Sachs, 1967). As C. Andrews (2021) points out, these claims pose a serious issue for cue-based retrieval models of memory, which often assume that syntactic features are available to the retrieval mechanism at significant distances, and sometimes even across sentence boundaries.

The field is a long way off from resolving this paradox. However, we can aim to make progress towards refining our theories of linguistic memory. There is a certain advantage to adopting a particular model of memory and seeing how far it can get us, and indeed, psycholinguistics has taken this tack with cue-based retrieval for over two decades, which has led to important findings about the nature of linguistic memory. Through this process, we've also seen that there are limitations to cue-based retrieval: there are linguistic phenomena that it alone cannot explain. The current work has made an attempt to adopt a relatively underexplored model within the sentence processing domain (see also Rich, 2024; Wagers, 2008), which shows promise in its explanatory potential. But, the nature of linguistic memory remains mysterious in many ways. Only by broadening our scope and integrating insights from elsewhere in memory research can we hope to push psycholinguistic theorizing forward.

References

AnderBois, S., Brasoveanu, A., & Henderson, R. (2010). Crossing the appositive/at-issue meaning boundary. In *Semantics and linguistic theory* (Vol. 20, pp. 328–346).

AnderBois, S., Brasoveanu, A., & Henderson, R. (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics*, *32*(1), 93–138.

Anderson, J. R. (1993). Production systems and the ACT-R theory. *Rules of the Mind*, 17–44.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*(4), 355.

Anderson, J. R., & Bower, G. H. (1974). Interference in memory for multiple contexts. *Memory & Cognition*, *2*(3), 509–514.

Anderson, J. R., & Paulson, R. (1977). Representation and retention of verbatim information. *Journal of Verbal Learning and Verbal Behavior*, *16*(4), 439–451.

Andrews, C. (2021). There and gone again: Syntactic structure in memory.

Andrews, S., & Veldre, A. (2021). Wrapping up sentence comprehension: The role of task demands and individual differences. *Scientific Studies of Reading*, *25*(2), 123–140.

Arvindam, V., & Wagers, M. (2022). Anti-local antecedents in Telugu are subject to local antecedent interference. *35th Conference on Human Sentence Processing*.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.

Astruc-Aguilera, L., & Nolan, F. (2007). Variation in the intonation of extra-sentential elements [Book Section]. In P. Prieto, J. Mascaro, & M.-J. Sole (Eds.), *Segmental and Prosodic Issues in Romance Phonology* (p. 85–107). John Benjamins.

Auran, C., & Loock, R. (2011). The prosody of discourse functions: The case of appositive relative clauses in spoken British English.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In *Reanalysis in Sentence Processing* (pp. 1–46). Springer.

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.

Barrouillet, P., & Camos, V. (2021). The time-based resource-sharing model of working memory. *Working Memory: State of the Science*, 85–115.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ...

Bolker, M. B. (2015). Package ‘lme4’. *Convergence*, *12*(1), 2.
Baumann, S., & Schumacher, P. B. (2020). The incremental processing of focus, givenness and prosodic prominence. *Glossa: a Journal of General Linguistics*, *5*(1).
Beckman, M. E., & Ayers, G. (1997). Guidelines for ToBI labelling. *The OSU Research Foundation*, *3*(30), 255–309.
Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in japanese and english. *Phonology*, *3*, 255–309.
Ben-Meir, N., Van Handel, N., & Wagers, M. (2019). Verbs retrieve subjects, not clausal attachment sites. *32nd Conference on Human Sentence Processing*.
Bever, T. G., Lackner, J., & Kirk, R. (1969). The underlying structures of sentences are the primary units of immediate speech processing. *Perception & Psychophysics*, *5*(4), 225–234.
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.
Blakemore, D. (2005). And-parentheticals. *Journal of Pragmatics*, *37*(8), 1165–1181.
Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze made easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language*, *111*, 104082.
Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, *67*(2), 135–144.
Breen, M., & Clifton, C. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, *64*(2), 153–170.
Brunetti, L. (2024). Contrast in a qud-based information-structure model. *On the Role of Contrast in Information Structure*, *382*, 191.
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*, 1–28.
Burton-Roberts, N. (1999). Language, linear precedence and parentheticals. *The Clause in English*, *45*, 33–51.
Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, *4*, 75–87.
Carlson, K., Frazier, L., & Clifton Jr, C. (2009). How prosody constrains comprehension: A limited effect of prosodic packaging. *Lingua*, *119*(7), 1066–1082.
Carlson, K., & Harris, J. (2018). Zero-adjective contrast in much-less ellipsis: The advantage for parallel syntax. *Language, Cognition and Neuroscience*, *33*(1), 77–97.
Caron, J., Micko, H. C., & Thüring, M. (1988). Conjunctions and the recall of composite sentences. *Journal of Memory and Language*, *27*(3), 309–323.
Carroll, J. M., & Tanenhaus, M. K. (1978). Functional clauses and sentence segmenta-

tion. *Journal of Speech and Hearing Research*, *21*(4), 793–808.
Chafe, W. L. (1979). The flow of thought and the flow of language. In *Discourse and Syntax* (pp. 159–181). Brill.
Chafe, W. L. (1980). The deployment of consciousness in the construction of narrative.
Chierchia, G., & McConnell-Ginet, S. (1990). Meaning and grammar: An introduction to semantics.
Chomsky, N. (1981). Lectures on binding and government. *Dordrecht: Foris.*
Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62.
Cinque, G. (2006). Two types of appositives. *University of Venice Working Papers in Linguistics*, *16*(1), 7–56.
Cinque, G. (2020). *The syntax of relative clauses: A unified analysis.* Cambridge University Press.
Cohen, H., Douaire, J., & Elsabbagh, M. (2001). The role of prosody in discourse processing. *Brain and Cognition*, *46*(1-2), 73–82.
Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49–59.
Dehé, N. (2009). Clausal parentheticals, intonational phrasing, and prosodic theory1. *Journal of Linguistics*, *45*(3), 569–615.
De Vries, M. (2005). Coordination and syntactic hierarchy. *Studia Linguistica*, *59*(1), 83–105.
De Vries, M. (2023). Are there different kinds of appositive relative clauses? *Linguistics in the Netherlands*, *40*(1), 230–247.
Dillon, B., Chow, W.-Y., Wagers, M., Guo, T., Liu, F., & Phillips, C. (2014). The structure-sensitivity of memory access: evidence from Mandarin Chinese. *Frontiers in Psychology*, *5*, 1025.
Dillon, B., Clifton Jr, C., 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. (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.
Dillon, B., & Wagers, M. (2019). Approaching gradience in acceptability with the tools of signal detection theory.
Drummond, A. (2010). *Ibex: Internet-based experiments.* `https://ibex.spellout.net/`. (Accessed: 2024-07-11)
Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., & Martey, N. (2000-2005). Santa Barbara Corpus of Spoken American English. *Philadelphia: Linguistic Data Consortium.*
Duff, J. (2023). *On the timing of decisions about meaning during incremental compre-*

hension (Unpublished doctoral dissertation). UC Santa Cruz.
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).
Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429–446.
Eerland, A., Engelen, J. A., & Zwaan, R. A. (2013). The influence of direct and indirect speech on mental representations. *PLoS One*, *8*(6), e65480.
Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior*, *21*(3), 296–306.
Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, *119*(2), 223.
Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, *30*(1), 3–20.
Féry, C. (2010). Recursion in prosodic structure. *Phonological Studies*, *13*, 51–60.
Fodor, J. (1998). Learning to parse? *Journal of Psycholinguistic Research*, *27*(2), 285–319.
Fodor, J. (2002a). Prosodic disambiguation in silent reading. In *North East Linguistics Society* (Vol. 32, p. 8).
Fodor, J. (2002b). Psycholinguistics cannot escape prosody. In *Speech Prosody.*
Fodor, J., Bever, A., Garrett, T., et al. (1974). The psychology of language: An introduction to psycholinguistics and generative grammar.
Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, *4*(5), 414–420.
Fodor, J. D., & Inoue, A. (1998). Attach anyway. *Reanalysis in Sentence Processing*, 101–141.
Foltz, P. (n.d.). *Word embedding analysis website.* `http://wordvec.colorado.edu/index.html`. (Accessed: 2024-07-03)
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.
Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, *41*, 163–171.
Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology*, *9*(7), S5–S22.
Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249.
Frazier, L., & Clifton, C. (1996). *Construal.* Mit Press.
Frazier, L., & Clifton, C. (1998). Sentence reanalysis, and visibility. In *Reanalysis in Sentence Processing* (pp. 143–176). Springer.
Frazier, L., Clifton, C., Carlson, K., & Harris, J. A. (2014). Standing alone with prosodic

help. *Language, Cognition and Neuroscience*, *29*(4), 459–469.
Frazier, L., Clifton, C., & Randall, J. (1983). Filling gaps: Decision principles and structure in sentence comprehension. *Cognition*, *13*(2), 187–222.
Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325.
Frey, W., & Truckenbrodt, H. (2015). Syntactic and prosodic integration and disintegration in peripheral adverbial clauses and in right dislocation/afterthought. *Syntactic Complexity across Interfaces*, *30*, 75.
Garrod, S. C., & Sanford, A. J. (1982). The mental representation of discourse in a focussed memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics*, *1*(1), 21–41.
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian Data Analysis*. Chapman and Hall/CRC.
Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331.
Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411.
Graesser, A., & Mandler, G. (1975). Recognition memory for the meaning and surface structure of sentences. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(3), 238.
Grice, H. P. (1975). Logic and conversation. In *Speech Acts* (pp. 41–58). Brill.
Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2017). bridgesampling: An R package for estimating normalizing constants. *arXiv preprint arXiv:1710.08162*.
Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, *12*(3), 175–204.
Gurguryan, L., Dutemple, E., & Sheldon, S. (2021). Conceptual similarity alters the impact of context shifts on temporal memory. *Memory*, *29*(1), 11–20.
Haberlandt, K., & Graesser, A. C. (1989). Processing of new arguments at clause boundaries. *Memory & Cognition*, *17*(2), 186–193.
Haberlandt, K., Graesser, A. C., Schneider, N. J., & Kiely, J. (1986). Effects of task and new arguments on word reading times. *Journal of Memory and Language*, *25*(3), 314–322.
Harrington Stack, C., & Watson, D. G. (2023). Pauses and parsing: Testing the role of prosodic chunking in sentence processing. *Languages*, *8*(3), 157.
Harris, J. (2016). Processing let alone coordination in silent reading. *Lingua*, *169*, 70–94.
Harris, J. (2023). The enduring effects of default focus in let alone ellipsis: Evidence from pupillometry. *Experiments in Linguistic Meaning*, *2*, 117–128.
Harris, J., & Carlson, K. (2016). Keep it local (and final): Remnant preferences in "let alone" ellipsis. *Quarterly Journal of Experimental Psychology*, *69*(7), 1278–1301.

Harris, J., & Carlson, K. (2018). Information structure preferences in focus-sensitive ellipsis: How defaults persist. *Language and Speech*, *61*(3), 480–512.
Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). *Detection Theory: A User's Guide*. Routledge.
Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, *26*(3), 699–720.
Hemforth, B., Fernandez, S., Clifton Jr, C., Frazier, L., Konieczny, L., & Walter, M. (2015). Relative clause attachment in german, english, spanish and french: Effects of position and length. *Lingua*, *166*, 43–64.
Hill, R., & Murray, W. S. (2000). Commas and spaces: Effects of punctuation on eye movements and sentence parsing. In *Reading as a Perceptual Process* (pp. 565–589). Elsevier.
Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, *88*(3), 297.
Hintzman, D. L., & Block, R. A. (1973). Memory for the spacing of repetitions. *Journal of Experimental Psychology*, *99*(1), 70.
Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Contextual associations and memory for serial position. *Journal of Experimental Psychology*, *97*(2), 220.
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.
Hirschberg, J., & Avesani, C. (1997). The role of prosody in disambiguating potentially ambiguous utterances in English and Italian [Book Section]. In *Proceedings of INT-1997* (p. 189-192). Athens, Greece.
Hitch, G. J. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 116–139.
Hoeks, M. (2023). *Comprehending focus/representing contrast* (Unpublished doctoral dissertation). University of California, Santa Cruz.
Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, *26*(3), 376–405.
Holler, A. (2008). A discourse-relational approach to continuation. *Pragmatics and Beyond*, *172*, 249.
Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.
Howard, M. W., Youker, T. E., & Venkatadass, V. S. (2008). The persistence of memory: Contiguity effects across hundreds of seconds. *Psychonomic Bulletin & Review*, *15*, 58–63.
Hulsey, S. M. (2008). *Focus sensitive coordination* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning*

& Memory, *14*(1-2), 47–53.
Hwang, H., Lieberman, M., Goad, H., & White, L. (2011). Syntactic ambiguity resolution: Effects of prosodic breaks and prosodic length. In *Proceedings of WCCFL* (Vol. 28, pp. 267–274).
Ishihara, S. (2022). On the (lack of) correspondence between syntactic clauses and intonational phrases. *Oxford Studies in Phonology and Phonetics*, 420–456.
Jackendoff, R., et al. (1977). *X syntax: A study of phrase structure.* MIT press.
Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 112.
Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 409–416.
Jarvella, R. J. (1973). Coreference and short-term memory for discourse. *Journal of Experimental Psychology*, *98*(2), 426.
Jarvella, R. J. (1979). Immediate memory and discourse processing. In *Psychology of Learning and Motivation* (Vol. 13, pp. 379–421). Elsevier.
Jasinskaja, K. (2016). Not at issue any more. *Ms. University of Cologne.*
Jun, S.-A., & Bishop, J. (2014). Implicit prosodic priming and autistic traits in relative clause attachment. In *Proceedings of the Conference on Speech Prosody.*
Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, *87*(4), 329.
Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228.
Kahana, M. J. (2012). *Foundations of Human Memory.* OUP USA.
Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative retrieval processes in episodic memory.
Kaland, C., & Heuven, V. J. v. (2010). The structure-prosody interface of restrictive and appositive relative clauses in dutch and german. In *Speech Prosody.*
Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, *56*(3), 384–409.
Keung, L.-C., & Staub, A. (2018). Variable agreement with coordinate subjects is not a form of agreement attraction. *Journal of Memory and Language*, *103*, 1–18.
Kim, N., Brehm, L., & Yoshida, M. (2019). The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval. *Language, Cognition and Neuroscience*, *34*(2), 190–213.
Kim, S., & Xiang, M. (2022). Memory retrieval selectively targets different discourse units. *35th Conference on Human Sentence Processing*.
Kim, S., & Xiang, M. (2023). Memory retrieval is sensitive to discourse status: through the lens of pronoun resolution. *36th Conference on Human Sentence Processing*.
Kim, S., & Xiang, M. (2024). Encoding discourse structure information during language

comprehension: Evidence from web-based visual world paradigm experiments. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46).

King, J., Andrews, C., & Wagers, M. (2012). Do reflexives always find a grammatical antecedent for themselves. In *25th annual Conference on Human Sentence Processing* (p. 67).

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194.

Klein, W., & Von Stechow, A. (1982). Intonation und bedeutung von fokus.

Klewitz, G., & Couper-Kuhlen, E. (1999). Quote–unquote? the role of prosody in the contextualization of reported speech sequences. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, *9*(4), 459–485.

Kluck, M., Ott, D., & De Vries, M. (2014). Incomplete parenthesis: an overview. *Parenthesis and ellipsis. Cross-linguistic and Theoretical Perspectives*, 1–21.

Koev, T. K. (2013). *Apposition and the structure of discourse.* Rutgers The State University of New Jersey-New Brunswick.

Krifka, M. (1992). *A compositional semantics for multiple focus constructions.* Springer.

Krifka, M. (1998). Additive particles under stress. In *Semantics and Linguistic Theory* (pp. 111–128).

Kroll, M. (2020). *Comprehending ellipsis* (Unpublished doctoral dissertation). UC Santa Cruz.

Kroll, M., & Wagers, M. (2019). Working memory resource allocation is not modulated by clausal discourse status. *Ms., University of California, Santa Cruz.*

Kush, D. (2013). *Respecting relations: Memory access and antecedent retrieval in incremental sentence processing* (Unpublished doctoral dissertation). University of Maryland, College Park.

Kush, D., Lidz, J., & Phillips, C. (2015). Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language*, *82*, 18–40.

Lascarides, A., & Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, *16*(5), 437–493.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course.* Cambridge University Press.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1787.

Lewis, R. L. (1998). Reanalysis and limited repair parsing: Leaping off the garden path. *Reanalysis in Sentence Processing*, 247–285.

Light, L. L., Stansbury, C., Rubin, C., & Linde, S. (1973). Memory for modality of presentation: Within-modality discrimination. *Memory & Cognition*, *1*(3), 395–400.

Liu, Y. S., & Caplan, J. B. (2020). Temporal grouping and direction of serial recall. *Memory & Cognition*, *48*, 1295–1315.

Logacev, P., & Vasishth, S. (2013). em2: A package for computing reading time measures for psycholinguistics. *R Package Version 0.9*.

Loock, R. (2007). Appositive relative clauses and their functions in discourse. *Journal of Pragmatics*, *39*(2), 336–362.

Lowder, M., & 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., Ryan, G., Opie, J., & Kaminsky, E. (2021). Effects of contrastive focus on lexical predictability during sentence reading: The case of not only... but also constructions. *Quarterly Journal of Experimental Psychology*, *74*(1), 179–186.

Magliano, J. P., Graesser, A. C., Eymard, L. A., Haberlandt, K., & Gholson, B. (1993). Locus of interpretive and inference processes during text comprehension: A comparison of gaze durations and word reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 704.

Manning, J. R., & Kahana, M. J. (2012). Interpreting semantic clustering effects in free recall. *Memory*, *20*(5), 511–517.

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science*, *189*(4198), 226–228.

Marslen-Wilson, W., & Tyler, L. K. (1976). Memory and levels of processing in a psycholinguistic context. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(2), 112.

McCarthy, J. J. (2011). *Doing optimality theory: Applying theory to data*. John Wiley & Sons.

McCawley, J. D. (1998). *The syntactic phenomena of english*. University of Chicago Press.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language*, *38*(2), 225–252.

McElree, B. (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*.

Mertzen, D., Laurinavichyute, A., Dillon, B. W., Engbert, R., & Vasishth, S. (2024). Crosslinguistic evidence against interference from extra-sentential distractors. *Journal of Memory and Language*, *137*, 104514.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, C. (2016). *Limited, syntactic reactivation in noun phrase ellipsis*. University of California, Santa Cruz.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our

capacity for processing information. *Psychological Review*, *63*(2), 81.
Miller, G., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, *2*(3), 217–228.
Miller, L., & Stine-Morrow, E. A. (1998). Aging and the effects of knowledge on on-line reading strategies. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *53*(4), P223–P233.
Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, *33*(1), 128–147.
Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269.
Nespor, M., & Vogel, I. (1986). *Prosodic Phonology.* De Gruyter Mouton.
Nevins, A., & Weisser, P. (2019). Closest conjunct agreement. *Annual Review of Linguistics*, *5*, 219–241.
Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411.
Oberauer, K. (2005). Control of the contents of working memory–a comparison of two paradigms and two age groups. *Journal of experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 714.
Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, *51*, 45–100.
Onea, E. (2016). *Potential questions at the semantics-pragmatics interface* (Vol. 33). Brill.
Ott, D. (2016). Ellipsis in appositives. *Glossa: a Journal of General Linguistics*, *1*(1).
Paape, D., & Vasishth, S. (2021). *When nothing goes right, go left: A large-scale evaluation of bidirectional self-paced reading.* PsyArXiv preprint. Retrieved from psyarxiv. com/d7pvz.
Paape, D., Vasishth, S., Paape, D., & Vasishth, S. (2022). Is reanalysis selective when regressions are consciously controlled? *Glossa Psycholinguistics*, *1*(1).
Pallier, C., New, B., & Bourgin, J. (2019). *OpenLexicon.*
Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, *157*, 321–339.
Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse.
Pizarro-Guevara, J. S., & Dillon, B. (2022). What tagalog can teach us: the influence of word order in reflexive processing. In *Proceedings of the 28th Meeting of the Austronesian Formal Linguistics Association.*
Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, *12*(5-6), 601–638.
Polyn, S. M., & Cutler, R. A. (2017). Retrieved-context models of memory search and the neural representation of time. *Current Opinion in Behavioral Sciences*, *17*, 203–210.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129.

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* (Vol. 7). OUP Oxford.

Pu, Y., Kong, X.-Z., Ranganath, C., & Melloni, L. (2022a). Event boundaries shape temporal organization of memory by resetting temporal context. *Nature Communications*, *13*(1), 622.

Pu, Y., Kong, X.-Z., Ranganath, C., & Melloni, L. (2022b). Event boundaries shape temporal organization of memory by resetting temporal context. *Nature Communications*, *13*(1), 622.

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 779.

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(4), 1061–1080.

Redeker, G. (2006). Discourse markers as attentional cues at discourse transitions. In *Approaches to discourse particles* (pp. 339–358). Brill.

Reeves, C., Schmauder, A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Rxperimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1638.

Reinhart, T. (1983). Coreference and bound anaphora: A restatement of the anaphora questions. *Linguistics and Philosophy*, 47–88.

Reinhart, T. (1991). Elliptic conjunctions-non-quantificational LF.

Repp, S. (2016). Contrast: Dissecting an elusive information-structural notion and its role in grammar.

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. *The Oxford Handbook of Metamemory*, *1*, 65–80.

Rich, S. K. (2024). *The features we use and the features we lose: encoding, maintenance, and retrieval* (Unpublished doctoral dissertation). UC Santa Cruz.

Riester, A. (2019). Constructing QUD trees. In *Questions in Discourse* (pp. 164–193). Brill.

Roberts, C. (1996/2012). Information structure in discourse: Toward a unified theory of formal pragmatics. *Ohio State University Working Papers in Linguistics*, *49*, 91–136.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 1–8.

Rooth, M. (1985). *Association with focus* (Unpublished doctoral dissertation). University of Massachusetts Amherst.

Rooth, M. (1992a). Ellipsis redundancy and reduction redundancy. In *Proceedings of the Stuttgart Ellipsis Workshop* (Vol. 29, pp. 1–26).
Rooth, M. (1992b). A theory of focus interpretation. *Natural Language Semantics*, *1*(1), 75–116.
Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*(9), 437–442.
Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory & Cognition*, *2*(1), 95–100.
Sahakyan, L., & Hendricks, H. E. (2012). Context change and retrieval difficulty in the list-before-last paradigm. *Memory & Cognition*, *40*, 844–860.
Schafer, A. (1997). *Prosodic parsing: The role of prosody in sentence comprehension.* University of Massachusetts Amherst.
Schafer, A., & Speer, S. (1998). Prosodic influences on the resolution of lexical ambiguity.
Schlenker, P. (2010). Supplements within a unidimensional semantics. In *Logic, Language and Meaning: 17th Amsterdam Colloquium* (pp. 74–83).
Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review*, *18*, 455–468.
Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893.
Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.
Selkirk, E. (2000). The interaction of constraints on prosodic phrasing. In *Prosody: Theory and experiment* (pp. 231–261). Springer.
Selkirk, E. (2011). The syntax-phonology interface. *The handbook of Phonological Theory*, *2*, 435–483.
Sharf, E. (2024). Pronoun interpretation and conceptual reactivation. *Ms., University of California, Santa Cruz*.
Simons, M., Tonhauser, J., Beaver, D., & Roberts, C. (2010). What projects and why. In *Semantics and linguistic theory* (pp. 309–327).
Simpson, H. E. (2016). *The role of intonation units in memory for spoken english.* University of California, Santa Barbara.
Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120.
Slowiaczek, M. L. (1981). *Prosodic units as language processing units* (Unpublished doctoral dissertation). University of Massachusetts Amherst.
Slowiaczek, M. L., & Clifton, C. J. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning and Verbal Behavior*, *19*(5), 573–582.
Smith, G., Franck, J., & Tabor, W. (2021). Encoding interference effects support self-

organized sentence processing. *Cognitive Psychology*, *124*, 101356.
Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. *Current Biology*, *27*(22), 3499–3504.
Speer, S. R., Crowder, R. G., & Thomas, L. M. (1993). Prosodic structure and sentence recognition. *Journal of Memory and Language*, *32*(3), 336–358.
Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25*(2), 249–271.
Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021766.
Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, *25*(5/6), 701–721.
Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in erps as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*(3), 267–295.
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.
Stowe, L. A., Kaan, E., Sabourin, L., & Taylor, R. C. (2018). The sentence wrap-up dogma. *Cognition*, *176*, 232–247.
Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*(2), 477–488.
Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, *11*(5), 882–888.
Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: a psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64.
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.
Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*(4), 355–370.
Tanenhaus, M. K., & Carroll, J. M. (1975). The clausal processing hierarchy... and nouniness. *Papers from the Parasession on Functionalism, Chicago Linguistic Society*, 499–512.
Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 37.

Toosarvandani, M. (2010). *Association with foci.* University of California, Berkeley.
Toosarvandani, M. (2013). Corrective but coordinates clauses not always but sometimes. *Natural Language & Linguistic Theory*, *31*, 827–863.
Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory and Language*, *59*(2), 223–236.
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 Handel, N. (2022). *The sound of silence: Investigations of implicit prosody.* University of California, Santa Cruz.
Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, *9*, 2.
Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... others (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*, 58–76.
Wagers, M. (2008). *The structure of memory meets memory for structure in linguistic cognition.* University of Maryland, College Park.
Wagers, M., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.
Wagers, M., & McElree, B. (2022). Memory for linguistic features and the focus of attention: evidence from the dynamics of agreement inside dp. *Language, Cognition, and Neuroscience*, *37*(9), 1191–1206.
Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory*, *28*, 183–237.
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.
Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production [Journal Article]. *Language and Cognitive Processes*, *19*(6), 713-755.
Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2018). Comprehension demands modulate re-reading, but not first-pass reading behavior. *Quarterly Journal of Experimental Psychology*, *71*(1), 198–210.
Westera, M. (2017). *Exhaustivity and intonation: A unified theory* (Unpublished doctoral dissertation). ILLC.
Westera, M. (2019). Rise-fall-rise as a marker of secondary QUDs [Book Section]. In D. Gutzmann & K. Turgay (Eds.), *Secondary content: The semantics and pragmatics of side issues* (p. 376-404). Brill.
Willer Gold, J., Arsenijević, B., Batinić, M., Becker, M., Čordalija, N., Kresić, M., ...

others (2018). When linearity prevails over hierarchy in syntax. *Proceedings of the National Academy of Sciences*, *115*(3), 495–500.
Willer-Gold, J., Arsenijević, B., Batinić, M., Čordalija, N., Kresić, M., Leko, N., ... others (2016). Conjunct agreement and gender in south slavic: From theory to experiments to theory. *Journal of the Slavic Linguistics Society*, 187–224.
Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research*, *41*, 105–128.
Wu, D. (2022). *Syntax and prosody of coordination* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, *23*(10), 3146–3152.
Yao, B., & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition*, *121*(3), 447–453.
Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX).* `https://doi.org/10.17605/OSF.IO/MD832`. (Accessed: 2024-07-03)
Zhou, P., & Christianson, K. (2016). Auditory perceptual simulation: Simulating speech rates or accents? *Acta Psychologica*, *168*, 85–90.
Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1196.
Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*(5), 292–297.
Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162.