Title
The Varying Roles of Morphosyntax in Memory and Sentence Processing: Retrieval and Encoding Interference in Brazilian Portuguese

Author
Lawn, Alexandra

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Varying Roles of Morphosyntax in Memory and Sentence Processing: Retrieval and Encoding Interference in Brazilian Portuguese

A dissertation submitted in partial satisfaction of

the requirements for the degree Doctor of Philosophy

in Hispanic Languages and Literatures

by

Alexandra Rae Lawn

2020

ABSTRACT OF THE DISSERTATION

The Varying Roles of Morphosyntax

in Memory and Sentence Processing:

Retrieval and Encoding Interference in Brazilian Portuguese

by

Alexandra Rae Lawn

Doctor of Philosophy in Hispanic Languages and Literatures

University of California, Los Angeles, 2020

Professor Antonio Carlos Quicoli, Co-Chair

Professor Jesse A. Harris, Co-Chair

Cue-based retrieval models have largely been adopted as a description of how linguistic content is retrieved from memory. Under this framework, a retrieval cue is projected at the site of a dependency and matched with its target using a parallel matching procedure (e.g., Van Dyke and Lewis, 2003). Although this is a highly efficient mechanism, retrieval difficulties occur when there are multiple items stored in memory that serve as potential matches for the retrieval cue(s), which is known as similarity-based interference (SBI). Several studies have demonstrated that a wide variety of linguistic information can generate SBI effects, but the theory of what serves as a retrieval cue is still relatively unknown (Van Dyke and Johns, 2012). Moreover, recent empirical evidence has proposed the similarity-based interference can arise from another source: the encoding mechanism (e.g., Villata et al., 2018).

Three hypotheses are addressed regarding three potential retrieval mechanisms: (1) a retrieval mechanism that only relies on cues relevant to the dependency being resolved, (2) one that is sensitive to all of the features overlapping between a target and distractor(s), or (3) a mechanism that is primarily sensitive to relevant features but produces additive interference effects for irrelevant features. Moreover, a fourth hypothesis investigates if similarity-based interference also arises from the encoding mechanism.

In an attempt to disentangle whether sentence processing disruptions occur as a result of retrieval mechanism (1) + encoding interference or due to one of the other mechanisms, 7 self-paced reading experiments were conducted on Brazilian Portuguese. In all of the studies, number was a relevant feature for the resolution of the grammatical dependency (subject-verb dependency in relative clauses or wh-remnant-correlate pairing in sluices) and gender features varied in their relevance. The rationale behind using these dependencies and features was to test whether syntactically relevant features produced stronger interference effects than irrelevant features and to propose why these results differed. Any findings that showed that irrelevant feature (gender) matches caused reading time slowdowns or decreased comprehension question accuracy before the retrieval site were interpreted as encoding interference.

Although results vary across studies, the findings in this thesis provide the most support for a combination of retrieval (mechanism 1) and encoding interference. Although the other two retrieval mechanisms cannot be completely ruled out at this time, the evidence that gender produces earlier and weaker effects reminiscent of encoding interference and that number produced interference reflective of retrieval interference are novel.

The dissertation of Alexandra Rae Lawn is approved.

Timothy Hunter

Ji Young Kim

Carson Schütze

Jesse A. Harris, Committee Co-Chair

Antonio C. Quicoli, Committee Co-Chair

University of California, Los Angeles

2020

To my grandma, Donna Bass

Thank you for always encouraging me to continue my education.

Thank you for inspiring my love of words and language.

Thank you for never letting me give up.

I miss you.

Table of Contents

List of Figures:

List of Tables:

Acknowledgments

As I look back at the journey that was graduate school, the proverb that *it takes a village* rings true. While I cannot thank every person that has crossed my path, inspired me, and encouraged me over the last five years, this would not have been possible without every single one of you.

I would like to begin by extending my endless gratitude to my committee co-chairs, Carlos Quicoli and Jesse Harris. Carlos brought me to UCLA. From the very first email I sent him while investigating graduate programs through the evolution of my dissertation project, his support has been unwavering. Thank you, Carlos, for going above and beyond to support my studies of Portuguese, linguistics, and for allowing me to explore my love of Brazil at all stages of this process. I will miss our meetings that covered everything from linguistics and research to soccer and Brazilian politics!

Jesse, you made me a psycholinguist. Thank you for your patience and willingness to train me from the ground up. Not only did you introduce me into the world of psycholinguistics, you taught me how important it is to participate in lab culture, explore my wildest research ideas (with proper research and pilot studies, of course!), and to be proud of the work that I produce. The lessons from your mentorship go far beyond academics. I will forever be grateful for everything that you taught me, the laughs in the lab, the endless professional guidance, and for truly showing me what it means to be scientist. I look forward to continuing our collaboration, especially with respect to the retrieval and encoding sluicing projects that are referred to in Chapter 4 of this dissertation.

Additionally, I want to thank Tim Hunter, Ji Young Kim, and Carson Schütze for being on my committee. Tim, thank you for your feedback, encouragement and time that you spent helping me produces this thesis. You have always encouraged me to investigate this encoding idea and that

really helped me keep going! Ji Young, you have always been an inspiration to me as a fellow experimentalist and taught me the importance of always considering sociolinguistic factors in my research. Carson, thank you for helping me think through logic, find confounds, and become a better experimentalist as a whole. You have all contributed significantly to my development as a researcher, and I can't thank you enough!

I also would like to thank two other professors put me on this path. James Hussar, thank you for being my mentor throughout my undergraduate degree and for pushing me to do a Ph.D. To this day, I model my classes and teaching after what I learned from you, and none of this would have been possible without you. Juan Ishikawa, thank you for being my first Portuguese professor and for giving me my first opportunity to be a TA for Portuguese. The Portuguese component of this dissertation wouldn't have happened without the knowledge of Portuguese and support that both of you gave me at CSUF.

To my friends, wherever you may be in the world right now, I don't know where I would be without you. First off, I want to thank my friends and colleagues that were part of my time UCLA. Luís Avilés González, Canaan Breiss, Jhonni Carr, Arturo Diaz, Marju Kaps, Briana Nettie, Iara Mantenuto, Gemma Repiso Puigdelliura, and Jeremy Steffman, you have all inspired me, pushed me, taught me, and been there for me during this process. I am so happy to have met all of you and can't wait to see the great things that you accomplish!

I also want to give a special shout-out to Adam Royer, Bethany Sturman, and Stephanie Rich. Adam, you have been my biggest supporter at UCLA (and outside of it!). You have taught me so much about language, life, the internet, and everything else. You have made me a better person, scholar, and friend. Beth, you are an inspiration. You are the hardest worker I have ever met, and you have the kindest soul I have ever seen. Thank you for inspiring me to be strong, kind

and independent in all of my pursuits. Stephanie, where do I even begin? You are one of the people that I look up to the most. We have been through so much together in this process, from research adventures while eating every product Trader Joe's has ever produced, to laughing and crying, and to just being downright weird. I couldn't have done this without you!

Também quero agradecer todos meus amigos e colegas no Brasil. Sem vocês e toda a ajuda que me deram, essa tese nunca teria sido possível. Primeiro, Thiago Motta Sampaio, obrigada por ter me dado a oportunidade de rodar meus experimentos no seu laboratório e por ter me ajudado tanto na Unicamp (no LAPROS). Continuando com o LAPROS, obrigada também a Fernando Sabatini pela ajuda durante os dois invernos que passei na Unicamp. Você realmente me ajudou a conquistar esse sonho!

E para minha segunda família: Marcell Barbosa Barros, Milton Neto, Victória Balassa, e Lucas Troise. Eu amo vocês. Não consigo agradecer vocês o suficiente por tudo. Passamos por tantas aventuras juntos e eu aprendi tanto com cada um de vocês. Milton, você é o irmão que o Brasil me deu. Você me ensinou a confiar, amar, e rir de uma maneira que eu nunca imaginei que seria possível. Ainda vamos dar o nosso rolê em Paris, dançando, cantando e rindo do nada.

I also need to take a moment to thank two extremely special people. Ino Samaras, you are my best friend. You have been there through thick and thin and your friendship has never wavered. You are the most selfless and genuine person I know, and life wouldn't be the same without you. Matthew Gallagher, you are my rock. You are incredibly loving, patient, and authentic, and I don't think I would have made it through this dissertation, pandemic, or quarantine without you. You inspire me to be better every day, and I look forward to staying weird with you.

Finally, to my Family. I love you all. You have been with me every step of the way, through every up and down, and every success and failure. Beth, your creativity, kindness, and daily phone

calls have kept me going when I didn't know I could. Katie, your wit, passion, and drive are truly an inspiration. Mom, thank you for always encouraging me to follow my own path, no matter how difficult it is. Thea, thank you for your advice, adventures and for feeding me when I'm too lazy to cook. Papa, your patience, kindness, and endless support make you the best dad that a daughter could ask for.

I am also grateful for the Mellon Pre-Dissertation grant, as well as UCLA graduate division and the Department of Spanish and Portuguese for providing the funding necessary to complete this thesis. Thank you everyone for all of your help and guidance, and any mistakes in this thesis are mine alone.

Thank you everyone. Muito obrigada.

VITA

Education

2020 (expected) — Ph.D. in Hispanic Languages and Literatures
University of California, Los Angeles

2016 — M.A. in Portuguese
University of California Los Angeles

2015 — B.A. in Spanish and Latin American Studies
State University of Campinas: Study Abroad, Spring 2013
University of Salamanca: Study Abroad, Spring 2012

Select Publications

2019 — Harris, Jesse A., **Alexandra R. Lawn** and Marju Kaps. *Investigating sound and structure in concert: A pupillometry study of relative clause attachment*. The 41st annual meeting of the Cognitive Science Society.

Select Presentations

2019 — **Lawn, Alexandra R.** & Jesse A. Harris. *Not all cues are equal: retrieving gender and number in Brazilian Portuguese sluiced sentences*. Poster presented at Going Romance, 2019. University of Leiden, Netherlands. November 28-29, 2019.

2019 — **Lawn, Alexandra R.** & Jesse A. Harris. *Similarity-based interference and morphological retrieval in Portuguese sluiced sentences*. Poster presented at the 32nd Annual CUNY Conference on Human Sentence Processing (CUNY). University of Colorado, Boulder, March 29–31, 2019.

2019 — Kaps, Marju, **Alexandra R. Lawn** & Jesse A. Harris. *Delayed attachment commitments for parenthetical relative clauses: An eye-tracking study*. Talk presented at the 32nd Annual CUNY Conference on Human Sentence Processing (CUNY). University of Colorado, Boulder, March 29–31, 2019.

2018 — **Lawn, Alexandra R.** and Jesse Harris. *Morphological retrieval and similarity-based interference in Portuguese sluiced sentences*. Talk presented at the 2nd California Annual Meeting on Psycholinguistics. University of Southern California. December 1-2, 2018.

2017 — **Lawn, Alexandra R.** and Jesse Harris. *Structural Bias, Cue-Relevancy and Similarity-Based Interference in Spanish Sluiced Sentences*. Talk presented at the 1st California Annual Meeting on Psycholinguistics. University of California, Los Angeles. December 2-3, 2017.

2016 **Lawn, Alexandra R.** *Gallego o galego? Atrapado en purgatorio lingüístico*. Talk presented at the XIII Spanish and Portuguese Graduate Student Conference. University of California, Los Angeles. April 21, 2016.

Research Grants & Fellowships

2019 Larry Lauerhass Graduate Summer Award
2019-2020 Dissertation Year Fellowship
2019 Mellon Pre-Dissertation
2019 Ben and Rue Pine Travel Grant
2018 Graduate Summer Research Mentorship
2017-2018 Graduate Research Mentorship
2017 Graduate Summer Research Mentorship

Chapter 1: Introduction

1.1 General Context: Memory and Sentence Processing

Upon receiving new linguistic information, the input must immediately be encoded into memory and stored long enough that the language processing mechanisms are able to identify and resolve the grammatical dependencies present in the structure, regardless of their complexity. Given that words are presented incrementally during online sentence processing and cannot be viewed in tandem, it is intuitive that the comprehender must temporarily retain all information into memory before they arrive at a final interpretation of the sentence. For example, in (1a), the subject *the graduate students* must be stored until the verb *studies* is reached, even though the constituents involved in the dependency are adjacent. More complex sentences, such as 1b-c, demonstrate an important fact about human language: grammatical dependencies, such as the relationship between a subject and a verb, can be separated by an arbitrary number of words and must maintain their relationship regardless of the quantity of intervening information. Operating under the assumption that all grammatical dependencies must be resolved, regardless of the amount of separation between related constituents, and the fact that memory is inevitably required to store information while the sentence is being processed, many questions have been raised in psycholinguistics about the relationship between the trans-domain human memory system and language processing system.

(1[1]) a. The graduate students study.
b. The graduate students in the coffee shop study daily.
c. The graduate students in the coffee shop with their colleagues study daily.

[1] The use of *studies* $_{\text{3RD.PL.PRESENT}}$ was intentional in these examples to parallel between Portuguese and English verb morphology. The relationship between number morphology and subject-verb agreement is a primary focus of this dissertation. The impoverished morphological system of English limits the cases in which number morphology is overt on the verb; however, this is always present in Portuguese – the language studied in this thesis.

Although the upcoming discussion of the involvement of memory mechanisms in sentence processing is by no means intended to replace traditional parsing procedures that initial phrase structure and syntactic interpretations for constituents (e.g. Frazier, 1987), a deeper understanding of how the memory system interacts with sentence processing can help us better understand the types of cognitive mechanisms that are involved in transforming a string of words into a coherent sentence in real time.

The idea that memory resources are a crucial component of sentence processing is not, in and of itself, a novel concept. In fact, discussions about the interaction between memory and sentence processing date back to the 1960s, when researchers were trying to understand why certain grammatical constructions were more challenging to process than other structures (e.g., Miller and Chomsky, 1963). Although proposals varied slightly, the underlying assumption of early memory-driven processing hypotheses was that working memory resources were limited and did not allow the comprehender to hold on to multiple unattached constituents while simultaneously processing incoming information (e.g. Miller and Chomsky, 1963). The inability to allocate memory resources to early-encoded items contributed to their decay and severely impacted their ability to be retrieved. Even though the concept of limited resources is highly intuitive when comparing sentences that vary in the number of open dependencies that they contain, it fails to completely explain why certain structurally identical sentences are not processed in the same manner. A growing body of psycholinguistic evidence has found that sentences like (2) tend to be easier to process than (3), but all that has changed between them is the semantic class of the relative clause subject (e.g., Gordon, 2001).

(2) The senator who Matthew disliked was voted out of office.

(3) The senator who lawyers disliked was voted out of office.

This example, like many that will be discussed in the following chapters, suggests that sentence processing is indeed affected by intervening information, but it is the content of this information, as opposed to its quantity, that has the greatest impact on comprehension. For the purpose of this discussion, linguistic *content* refers to the features associated with words in a sentence (e.g., gender, number, Case). In (2-3), stating that the linguistic content affected sentence comprehension refers to the fact that the presence of a proper name in (2) facilitated sentence comprehension when compared with a nearly identical structure containing two common nouns. Using just this information, one might draw the conclusion is that sentences containing words with *similar* linguistic qualities incur greatest processing penalties than those where the content is unique.

1.2 Motivation for the current studies

The human memory system, regardless of the cognitive domain that requires its usage (e.g., language comprehension, visual memory, number recall), shows sensitivity to overlapping information (e.g., Van Dyke and Johns, 2012; Makovski et al., 2009; McElree, 2001). In general, the greater the degree of overlap between multiple items stored in memory, the more challenging it becomes to retrieve the intended target. In psycholinguistics, these findings have inspired the development of the cue-based retrieval models. In this framework, linguistic information is retrieved from memory through a cue-matching procedure that occurs between a probe containing a *retrieval cue* and a *target* that overlaps with this feature. As seen in other cognitive tasks, when multiple items have been encoded with the same linguistic features, the probe becomes *cue-overloaded* and cannot easily identify the intended target from its competitors (Watkins and Watkins, 1975). This overload leads to processing penalties in both online and offline psycholinguistic studies, which are oftentimes referred to as *similarity-based interference* effects.

Although cue-based retrieval has a substantial amount of support in the literature (see Parker et al., 2017 and Van Dyke and Johns, 2013 for overviews), a significant amount of research is still needed to identify the nature of the retrieval mechanism, the linguistic information involved in retrieval, how linguistic content becomes available to the retrieval mechanism, and the precise mechanism driving similarity-based interference effects.

In this dissertation, I will be focusing on four current problems in the theory: (1) the lack of consensus about how the retrieval mechanism behaves when more than one cue is involved, (2) how independently motivated processing biases interact affect retrieval, (3) whether all interference effects arise during retrieval or if there is a second mechanism that causes these effects to surface during the encoding stage, and (4) how the features that are irrelevant to a dependency modulate interference. These issues do not constitute the complete list of problems that have been raised about the implementation of cue-based retrieval models in processing theories, but they make significant headway in addressing some of the core issues that must be considered in the development of this framework. Aside from these theoretical issues, there has been very little research conducted on retrieval in non-English languages, which is necessary to support the claims that the retrieval mechanism exists for all languages.

The experiments in this thesis attempt to overcome some of these gaps by testing the involvement of morphological features in distinct long-distance dependencies found in relative clauses and sluices in Brazilian Portuguese. Although memory is inherently involved in processing both long-distance and adjacent dependencies, from here on out, I focus on long-distance dependencies to more directly address questions related to similarity-based interference. as interference effects surface primarily when distractors are encoded into memory. Since long-distance dependencies allow for the insertion of distractors that have been shown to compete with

the intended target of the retrieval cue, they present a controllable testing ground to examine the limitations of memory with respect to language and how these limitations may influence online sentence comprehension.

1.3 Overview of key theoretical concepts and terminology

This section provides a brief overview of the theoretical concepts and terminology that will be referenced throughout this dissertation. This is by no means intended to be a comprehensive review of the literature, and an in-depth discussion of these topics can be found in Chapter 2. These descriptions are intended to present the context behind the research questions that are outlined in Section 3, as well as provide the general motivation behind the experiments conducted in the subsequent chapters.

1.3.1 Cue-based retrieval

As the sentences in example (1) suggest, grammatical dependencies can be separated by an arbitrary number of words, but regardless of distance, all dependencies must be resolved for the proper interpretation to be obtained. The resolution of grammatical dependencies is largely syntactic in nature, likely resulting from an initial structure-building stage, but certain dependencies (e.g. pronouns) make it clear that a retrieval mechanism is still required to retrieve antecedents from memory before a final interpretation can be obtained. In this dissertation, I will be focusing primarily on how antecedents are retrieved from memory, as well as how they are encoded. On the surface, the operations behind linguistic retrieval are quite straightforward. Upon arriving at a dependent element (e.g., a verb), the reader encounters retrieval cue that provides to direct accesses all potential targets for the given the dependency (e.g. nouns that are subject-like). This retrieval cue acts as a "signal" during online processing that the constituent actively being processed requires an antecedent that has been previously encoded, although the syntactic

relationship between the dependent elements might have already been grammatically parsed in certain constructions. The exact temporal ordering between the initial parse and retrieval is not clear, but there is some evidence that structural features, such as Case, can generate in interference. This implies that, in some cases, parsing logically precedes retrieval. implies that some degree of parsing must precede the retrieval of an antecedent. However, as examples (2-3) demonstrate, the ease of retrieving a target is significantly affected by the amount of overlap between the items that have been stored in memory. The greater the degree of overlap between competitors and the target (e.g., multiple nouns of the same class in example 3), the more challenging it becomes for the cue to match with the correct target. This concept is not exclusive to language processing and is oftentimes referred to as the fan effect, which has been observed in a wide variety of non-linguistic tasks (see Anderson, 2004 for a discussion of the fan effect). This topic will be explored in Chapter 2 when several models of retrieval are presented.

The *cue-based retrieval* framework was designed to account for the observations above. In these models, all potential antecedents for a given retrieval cue are accessed in parallel through a rapid, associative cue-matching procedure (see Chapter 2 for discussion). This matching procedure occurs when a retrieval *cue* at the site of a dependency is associated with the content of an item that has been previously encoded in memory. Example (4) demonstrates how cue-based retrieval operates on an abstract level. In (4), we observe that the verb contains retrieval cues for a nominative, animate NP and is paired with the verb without interference from a distractor NP. It is also important to note that the NP *the graduate students* contains a variety of other features that are not necessarily matched with the retrieval cue, implying that the cue-based retrieval mechanism only focuses on a subset of features that can be directly matched between the target and cue. The

behavior of these extraneous features will be a crucial part of the research presented in this dissertation.

Unfortunately, the exact definition of a *cue* is still unclear, but there is evidence that a linguistic retrieval cue is composed of the grammatical, contextual and structural information associated with a specific word, which, when combined, form a subset of the features of the intended antecedent (Lewis and Vasishth, 2005; overview in Van Dyke and Johns, 2012). For the sake of this dissertation, I treat linguistic *features* to be the content by which retrieval cues and targets are matched, such as those provided in the matrix underneath *graduate students* in (1). The concept of features in linguistic research can vary greatly depending on the theoretical framework, but in the upcoming experiments, I will be focusing primarily on morphosyntactic features that can be overtly (gender, number) or covertly expressed (Case).

(4) The graduate students study.

$$\begin{bmatrix} +\text{NP} \\ +\text{Nominative} \\ +\text{Plural} \\ +\text{Animate} \end{bmatrix} \qquad \begin{bmatrix} seeking + \text{NP} \\ seeking + \text{Plural} \\ seeking + \text{Nominative} \\ seeking + \text{Animate} \end{bmatrix}$$

This last factor raises an important challenge for models of cue-based retrieval: how are underspecified or covert features treated in retrieval? One could imagine that example (4) is changed to *the graduate student(s) studied*, where the English past tense does not contain an overt number feature. In cases like this, there is even stronger evidence that the string has been parsed on some level so that minimally, certain structural features could be established to help resolve the morphosyntactic ambiguity associated with words like *studied*. While this is a very important topic, one of the reasons that Portuguese was selected for this dissertation is because

morphosyntactic features like number and gender are always *overtly* expressed, which allows the discussion about covert features to be left for future research. I do not intend to claim that lexically transparent features are the only features employed in retrieval, as Wagers and Phillips have pointed out (2009, 2013), nor was this dissertation intended to determine or generate a set of potential features involved in retrieval. Rather, I employ lexical features that are known to generate interference effects to test how they behave when they are either directly or indirectly involved in the resolution of a grammatical dependency.

Another crucial component in cue-based retrieval is that the models are content-addressable (e.g. Van Dyke and Lewis, 2003; McElree, 2001; Martin and McElree, 2008). Content addressability, with respect to sentence comprehension, has been defined as a component of the memory mechanism where "syntactic and semantic constraints provide direct access to relevant representations without the need to search through potentially irrelevant information" (McElree, 2000). This aspect of the cue-based retrieval model is important, as it generates the prediction that the quantity of information stored in memory will not impact sentence processing speed, so long as the information does not overlap with the linguistic content involved in the [retrieval cue + target] pairing procedure. Content addressability is crucial for efficient retrieval as a retrieval cue associated with a verb that requires a subject will never undergo a cue-matching procedure with an extraneous syntactic element, such as another verb. Instead, the retrieval cue will have direct and parallel access to all potential targets (i.e., nouns in subject position) that at least partially overlap with the retrieval cue. This component of the model is one of the ways in which the potential set of targets for a retrieval cue is constrained, a topic which will be referenced frequently in the upcoming chapters.

Finally, cue-based retrieval models appear to be parallel-search systems, which also increases their efficiency (Lewis and Vasishth, 2005; Van Dyke and Johns, 2012 for overview). Evidence for the parallel nature of retrieval has been obtained from Speed Accuracy Tradeoff paradigms, an experimental method that measures the time it takes for comprehenders to obtain a stable interpretation of the information being presented to them (McElree and Dosher, 1989; McElree 2000, 2006 among others). In these tasks, comprehenders make binary grammaticality judgments at predetermined intervals at various points throughout the sentence and following. This information is then used to measure when and where a stable interpretation of the sentence has been reached (and the sentence is deemed acceptable). These studies have shown that only retrieval accuracy decreases when processing long-distance grammatical dependencies, but the time it takes to reach a stable interpretation does not (see Foraker and McElree, 2011 for an overview of experiments). These results suggest that retrieval is not serial in nature but rather a parallel matching procedure that is sensitive to the competition provoked by storing multiple representations in memory that have overlapping content. In a sentence processing context, these findings imply that the linear distance between grammatical dependencies is not the primary source of processing delays since the retrieval mechanism does not access items serially. What is important, however, is the amount of content overlap between items that are accessed during the retrieval process.

1.4 Similarity-Based Interference in Retrieval

If content-addressable, cue-based retrieval is so efficient, how can it explain sentence processing difficulties? Several studies have shown that the cue-based retrieval mechanism is extremely sensitive to information overlap between items stored in memory (Van Dyke and Lewis, 2003; Lewis and Vasishth, 2005, among others). Retrieval becomes more challenging in cases where the

retrieval cue could match with more than one item in a sentence (e.g. Van Dyke and Lewis, 2003; Lewis and Vasishth, 2005; aligns with fan effect, Anderson, 2004). The concept that information becomes harder to recall as more information is learned is not exclusive to retrieval models and is oftentimes described as the *fan effect* in the psychological literature (e.g., Anderson, 2004). Returning to examples (2-3) in section 1.1, example (3) is harder to process because the retrieval cue *the judge* and *the senator* belong to the same semantic class, which does not occur in (2), where one noun is a common noun and the other is a proper name. The increased semantic overlap in (2) makes it harder for the retrieval cue projected by the matrix verb to match with the intended target (the matrix subject, *the senator*), resulting in slower and less accurate retrieval. The inability of the retrieval cue to unambiguously identify its target results in *similarity-based interference*, which tends to surface as processing delays in online studies and inaccurate comprehension question responses in offline studies (Van Dyke and Johns, 2012 for an overview).

Similarity-based interference can be further categorized as being either *proactive* or *retroactive*, depending on the encoding order of the target and the distractor. Proactive interference occurs when the intended target is placed between a distractor and the retrieval cue (Distractor – Target – Probe), whereas *retroactive interference* occurs when the distractor is placed between the target and the probe (Target – Distractor – Probe). In sentence processing, both retroactive and proactive interference have been observed (*Retroactive*: Harris, 2015, 2019; McElree et al., 2003; McElree, 2009, Van Dyke and McElree, 2011, Sasson, 1971, etc.; *Proactive*: Fedorenko et al., 2006; Gordon, 2002; Van Dyke and McElree, 2006). It is important to note, however, is that McElree et al., (2003) found that retroactive interference may play a larger role in sentence processing than proactive interference.

The presence of similarity-based interference, or even its absence when it is predicted to occur, is the primary method that has been used to identify which linguistic features act as retrieval cues. In this dissertation, I follow this approach by testing whether number and gender features are able to produce similarity-based interference effects in relative clauses and sluices. Although several studies have tested the effects of number as a source of similarity-based interference (e.g. Harris, 2015, 2019; Wagers et al., 2009), gender features have been less studied due to their limited grammatical usage cross-linguistically (e.g. half of the documented languages on WALS do not contain gender systems; Corbett, 2005). Although languages such as Portuguese and English both rely on number information to a certain extent in subject-verb agreement, gender agreement between nouns is a grammatical component of Brazilian Portuguese that is not required by a verb seeking a subject. Since gender is not required in this dependency, this dissertation specifically compares its behavior in retrieval scenarios to that of number in order to address the theoretical gaps discussed in section 1.2

1.5 Encoding as a source of processing difficulty

As mentioned in the introduction, sentence comprehension requires information to be encoded into memory before it can be retrieved. While the process of retrieving information and the effects of similarity-based interference are relatively well-attested in the psycholinguistic literature, the mechanism(s) responsible for encoding have received less attention. Logically, if there is evidence that the human memory system is sensitive to information overlap during retrieval, it is possible that these effects also impact how we store information, to begin with. In other words, it is very challenging to determine if similarity-based interference effects are purely retrieval driven or if they can also stem from processes that precede the involvement of the retrieval mechanism, such as encoding.

Unfortunately, the literature on encoding interference is scarce, and the findings differ between studies. From a theoretical perspective, it has been claimed that encoding interference occurs as a result of a process known as *feature overwriting* (e.g., Nairne, 1960). Feature overwriting occurs when items (A) and (B) are encoded into memory with the same feature, where (A) is linearly encoded before (B). Given the inability of the memory system to allow the same feature to be equally activated across multiple items, the activation level of the overlapping feature in (A) is decreased, while the activation levels of the same feature in (B) are unaffected.[2] This process creates a type of similarity-based interference between items (A) and (B) where (A) is now less active in memory than (B), whereby becoming harder to retrieve.

Example (5) provides a concrete linguistic example of how this process functions. To begin, NP1 *as meninas* (the$_{\text{FEM.PL.}}$ girls$_{\text{FEM.PL.}}$) is encoded before NP2, *a atriz* (the$_{\text{-FEM.SG.}}$ actress$_{\text{FEM.SG.}}$) and both NPs share feminine gender.

(5) As meninas que adoram a atriz...
The$_{\text{-FEM.PL.}}$ girls$_{\text{-FEM.PL.}}$ that adore the$_{\text{-FEM.PL.}}$ actress$_{\text{-FEM.PL.}}$ ...
The girls that adored the actress...

...assistiram uma entrevista com ela.
...watched$_{\text{-3rd.PL.}}$ an interview with her.
...watched an interview with her

The memory mechanism cannot store both feminine nouns equally, so feature overwriting occurs. This implies that *the girls* is now less active than *the actress*, which can complicate the ability for the retrieval cues projected by the matrix verb *watched* to retrieve the intended subject. Although the retrieval cue does not require a gender match between the cue and target, feature overwrite of the gender feature could indeed generate similarity-based interference since the overall activation

[2] The degree of activation level changes varies by model. For example, Nairne (1960) suggests complete overwrite, as I have described here, but Villata et al. (2018) state that the activation levels of (A) and (B) become equal by lowering them in (A) and raising them in (B).

levels of the target and distractor have been altered by this process. What makes this mechanism challenging to identify and support in empirical studies, however, is that the features that overlap during encoding are oftentimes also involved in the retrieval process, there is no way to tell which mechanism causes interference effects in most online studies.

Although some current models address how encoding interference arises, feature overwriting being the most referenced, there are several empirical challenges to overcome before encoding interference can be isolated from retrieval interference in current retrieval models. This is because encoding interference can only be truly separated from retrieval interference if it appears on features that are not expected to generate retrieval interference under a purely cue-based model of sentence processing. In fact, Lewis and Vasishth (2005) point out that an appropriate account of encoding interference is missing in nearly all current processing models and could actually explain some of the evidence for similarity-driven processing difficulties that occur before retrieval sites (e.g., Gordon et al., 2004). Fortunately, in languages with complex morphological systems, it becomes possible to test for encoding effects from features that overlap due to grammatical constraints but are not directly involved in retrieval. For example, Villata et al. (2018) conducted a study on Italian relative clauses. As with Portuguese, the language studied in this dissertation, Italian subject-verb dependencies only require number agreement, but all nouns additionally express gender morphology that is unrelated to the dependency. The authors tested if gender produced any type of processing penalty despite its lack of involvement in the dependency. They found weak online effects for gender matches in relative clauses and strong offline effects. Since gender was not involved in the dependency, they interpreted the effects as evidence for encoding interference. In addition, they point out that although several have failed to find these online effects for encoding interference (e.g., Jäger, 2015), all of the reported studies on the topic

have found that encoding features decrease offline sentence comprehension. In this thesis, I conduct a set of studies on Portuguese relative clauses and sluices in order to complement the findings from Villata et al. (2018), as well as expand their claims to a new set of structures.

1.6 Research Questions

Although there are numerous studies showing that similarity-based interference and retrieval play an integral role in language comprehension, there are still many open questions regarding similarity-based interference and the mechanisms that generate this effect in sentence processing. To begin, I address the fact that the majority of the research on retrieval has focused on the behavior of one *relevant* linguistic feature that is paired with a retrieval cue within a specific syntactic context (e.g., semantic class overlap in relative clause processing, Gordon, 2001). A relevant feature, throughout this dissertation, refers to an overt linguistic feature that is required to resolve a grammatical dependency, although covert features may still be involved and merit further research. The experiments in this dissertation focus on the effects of morphosyntactic features which can either be *relevant* to a dependency (e.g., the effect of number features in subject-verb agreement) or *irrelevant* to a dependency, meaning that they are grammatically expressed by the language but are not overtly expressed at the retrieval site (e.g. NP-internal concord that does not influence subject-verb agreement). This is not to say that covert features, such as Case, are not involved in the retrieval process, and these features should also be tested for interference effects in languages that overtly express them to deepen our understanding of their behavior in other languages. The example below from Portuguese demonstrates this distinction between relevant (number) and irrelevant (gender) features in subject-verb dependencies, where NP internal gender agreement is grammatically required but not relevant in the retrieval process because it does not

serve to identify a target from its distractors in subject-verb dependencies (e.g. *comeram* "ate" the example below),

(6) Os gatos gordos comeram meu jantar inteiro.
The$_{\text{-M.PL}}$ cats$_{\text{-M.PL.}}$ fat$_{\text{-M.PL.}}$ ate$_{\text{-PL}}$ my$_{\text{-M.SG.}}$ dinner$_{\text{-M.SG.}}$ whole$_{\text{.SG.}}$
The fat cats ate my whole dinner

The concept of relevancy is particularly important, as it can produce a more realistic theory of the relationship between retrieval cues and the sets of features that are encoded alongside words in memory. As mentioned in the introduction on cue-based retrieval, a retrieval cue matches with a relevant *subset* of features on its intended target, but it is still unclear in the theory whether the remaining encoded features influence sentence processing. Moreover, by identifying that only linguistic features relevant to the dependency in question are involved in retrieval-driven interference, we can take steps towards limiting the exact linguistic content that is considered by the memory system in retrieval procedures. This leads us to the first, and primary research question of this dissertation:

> (1) How does the retrieval mechanism treat relevant and irrelevant features that have been encoded on constituents/words previously encoded in memory?
>
> (1a) Are *all* features capable of modulating similarity-based interference effects or are there crucial differences in the behavior of features with respect to their relevancy in the retrieval process?

Moving forward, we arrive at another issue in cue-based retrieval theory: handling multiple cues. Although identifying the behavior of a single feature in retrieval is beneficial for developing a coherent theory of what constitutes a cue in language processing, this is not necessarily reflective of natural language. If one were to assume that any linguistic feature relevant to the resolution of a grammatical dependency can be involved in retrieval, then there are cases where more than one cue is involved in the process. Although there has been relatively recent work on the combinatorics of multiple cues, it is still very unclear how the retrieval mechanism processes multiple cue overlap

(Parker et al., 2017). If distinct linguistic information shows varying effects on retrieval (e.g., gender features cause more interference effects than number features), we have suggestive evidence that the retrieval mechanism may assign weights to certain cues. There is also experimental evidence that the degree of informativity of a cue (diagnosticity) modulates the time-course in which SBI effects appear (Harris, 2015, 2019). This finding alone supports that not all cues are predicted to generate the same interference profile (also referenced in Martin, 2016) We must then speculate why certain cues are more active in retrieval than others and test if these hypotheses hold in a variety of contexts.

> (2) When more than one feature is relevant to the dependency, how does the retrieval mechanism behave?
>
> (2a) Do multiple overlapping features result in additive or multiplicative processing delays, or does retrieval appear to be driven by only one feature at a time?
>
> (2b) If features are indeed weighted, what generates these weights? Some options that will be explored are dialectal preferences, statistical biases, and structure-specific weights.

While the first two research questions will allow for more concrete definitions of how the retrieval mechanism treats irrelevant and relevant features, as well as how these features might be weighted in certain contexts, they do not directly address the mechanisms that are driving these effects. In particular, current retrieval models fail to predict processing delays for irrelevant features at both the retrieval and encoding levels (Villata et al., 2018; Jäger et al., 2015, 2017). Since irrelevant features are grammatically required and alter the overall meaning of the sentence, it is highly probable that they are encoded alongside the relevant retrieval features. Franck and Wagers (2020) point out that these "irrelevant features" (e.g. gender in Romance subject-verb dependencies) should be the focus of research investigating the potential of encoding interference on sentence processing. Unfortunately, a very limited set of models have been proposed to account

for interference effects from irrelevant features in a way that aligns with our current understanding of retrieval (e.g., Oberauer and Kliegel, 2006; Villata et al., 2018). Therefore, the third primary research question addresses the possibility of an encoding mechanism, alongside a retrieval mechanism, that can modulate interference effects in sentence processing.

> (3) Assuming that relevant and irrelevant features both produce interference, can these effects best be explained through the incorporation of an encoding mechanism into current models of memory and sentence processing?

In other words, can the retrieval-only models successfully account for the behavior of relevant and irrelevant features or is there evidence that a pre-retrieval process provides more explanatory power? This question, although it sits at the very heart of this dissertation, is presented last, as addressing it assumes that there will be a difference between irrelevant and relevant features that cannot be explained by pre-existing retrieval models alone. Therefore, the discussion of this specific question will correspond to the empirical discussions and findings in Chapters 3 and 4. Moreover, proposing a theory that now possesses both an encoding and retrieval mechanism is inherently more complex than a system that operates using a retrieval-only approach (with proper modifications). While a more complex model may provide a greater degree of explanatory power, the burden of proof is much greater than simply modifying a retrieval mechanism that could potentially account for the same effects. Given the difficulty of finding encoding effects in the literature (see Jäger et al., 2015 for a discussion), as well as the empirical inconsistencies that will be presented in Chapters 3 and 4, this question opens the door to theoretical speculations that could be treated as amendments to retrieval current models and encourage further developments and research regarding encoding and sentence processing.

The experiments conducted in this dissertation were designed to address all three of these questions, to varying extents. In Chapters 3 and 4, I address the behavior of gender and number

morphology in Portuguese sluices and relative clauses. These two structures have been shown to contain specific processing biases which induce similarity-based interference effects when violation (see Harris, 2015, 2019 for sluices; Gordon et al., 2001 for relative clauses). Moreover, while number is always a relevant feature in both dependency types, sluices allow for the manipulation of gender relevancy. In Portuguese sluices, it is possible for both gender and number morphology to be expressed on the remnant, which makes them both relevant features for retrieval. The expression of gender on the wh-remnant, however, is optional on wh-remnants, which allows for the direct manipulation of gender as both a relevant and irrelevant feature. This optionality allows for a direct comparison between the behavior of a single feature in retrieval (number) and two relevant features (gender and number). A comparison of this nature, to the best of my knowledge, has not been conducted in a controlled fashion, and any results will be highly insightful.

Given the ability to manipulate gender as an irrelevant feature in relative clauses and sluices, these experiments also allow me to directly address encoding interference and also helps to narrow down the relationship between feature relevancy and retrieval cues. Although we have good reason to assume that the linguistic features that translate to retrieval cues are only those that are required by a grammatical dependency, this dissertation will manipulate the relevance of gender to test for any potential involvement in retrieval and/or encoding interference. If I were to find strong effects of gender interference when the cue was not relevant, there are two possible interpretations: (1) the retrieval mechanism is also sensitive to features that overlap but are not involved in the dependency or (2) interference generated by irrelevant feature matches are evidence for encoding interference. These two possible interpretations will be addressed in relation

to Brazilian Portuguese, and although these studies cannot provide a concrete response to these questions, it makes significant steps towards disentangling these two approaches.

1.7 Chapter overview

Following this introduction, there are four remaining chapters in this dissertation that aim to address the research questions above. Chapter 2 presents a broad overview of both the psychological and psycholinguistic literature on the involvement of memory in sentence processing. After presenting the literature in support of cue-based retrieval as a means of resolving grammatical dependencies, I address the topics of similarity-based interference in both encoding and retrieval. To wrap up, I present the general hypotheses that are to be addressed throughout the dissertation, as well as the empirical predictions that they generate.

Chapters 3 and 4 contain the empirical studies that were conducted to test for retrieval and encoding interference effects in Brazilian Portuguese. In both chapters, the morphosyntactic features of gender and number are manipulated. Chapter 4 examines an ellipsis construction known as sluicing. As in standard retrieval procedures, like those found in subject-verb dependencies, ellipsis requires the remnant of an elided clause to be paired with an overtly expressed correlate. This remnant-correlate pairing procedure is similar to standard cases of retrieval (e.g. subject-verb dependencies) and shows sensitivity to similarity-based interference. A set of studies by Harris (2015, 2019) focused on the cues that allow the processor to find an appropriate correlate-remnant pairing and found retroactive interference effects (Harris, 2015, 2019). From a slightly different perspective, Martin and McElree (2008) focused on the content addressable nature of the retrieval mechanism and found that the interference present in ellipsis aligns with our current understanding of content-addressability. Aside from attempting to replicate studies on English sluicing that demonstrate that the diagnosticity of cues and processing biases affect how these sentences are

processed (e.g. Harris 2015, 2019), Brazilian Portuguese sluices allow for a more flexibility in the type of cues that can be tested, as gender can be either a relevant or an irrelevant feature. In the majority of the studies in this dissertation, gender is not required by a dependency, which makes its role in retrieval less straightforward. However, being able to show that gender is indeed a functional retrieval cue in appropriate contexts allows for a direct comparison between its behavior as both a relevant and irrelevant feature, which will ultimately help to address the hypotheses presented in Chapter 2.

In Chapter 3, Object- and Subject-Extracted relative clauses were tested for interference effects using self-paced reading (ORCs and SRCs, respectively). In the design of the experiments presented in the chapter, I relied on the overwhelming body of literature indicating that ORCs are harder to process than SRCs, as well as several studies that show that minimizing the degree of interference between the matrix and relative clause NPs reduces the asymmetry. Any interaction between Clause Type and Feature Match (number or gender), showing that feature matches in ORCs receive the greatest processing penalty are interpreted as evidence for retrieval-based interference. In addition, any main effect of feature match that is unrelated to the clause type is potential evidence for encoding interference, as the literature only predicts that feature matches will cause retroactive retrieval interference effects.

Chapter 5 contains a general discussion of the findings in the empirical chapters. Although the structures addressed in Chapters 3 and 4 are quite distinct and generate different processing patterns, the hypotheses behind the studies remain the same. To further motivate why these two constructions were used, they both contain well-known processing asymmetries that are driven by the position of the target (object-extraction makes RCs harder to process, Gordon et al., 2001; subject-correlates make sluices harder to process, Harris 2015, 2019). Previous claims that these

asymmetries are capable of modulating interference effects during online processing opens the door for a discussion about the potential interaction between features involved in retrieval and structural biases/preferences. Conveniently, both constructions use the same overt morphological features in the resolution of their dependencies, which allows for a direct comparison of gender and number features in Portuguese sentence processing. As I will later discuss, the experimental findings differ slightly between the relative clauses and sluices, but there are very important trends found between the two constructions with respect to retrieval and encoding.

Overall, the objective of this dissertation is to further refine the theory of cue-based retrieval by testing several hypotheses about how the mechanism works. The hypotheses put forth in Chapter 2 present three alternatives for the structure of the retrieval mechanism: one that (1) only operates using only features relevant to the dependency, (2) uses *all* features that have been encoded into memory or (3) treats multiple features as super-additive or multiplicative. Moreover, I address the concept of *relevancy*. I examine whether features that are required (relevant) by a dependency show unique behavior when compared to features that are not required, which I use to narrow down the retrieval hypotheses. Finally, I explore the concept of encoding interference, which is where this dissertation makes the most theoretical contributions. Although none of the findings in this study can conclusively say that encoding interference is affecting the retrieval process, they provide very strong evidence in favor of such a claim.

Chapter 2. Review of the Memory Literature:

2.1 What is the role of memory in sentence processing?

The review of the literature in this chapter describes the evolution and current state of content-addressable cue-based retrieval models, which focus on the role of memory and informational distinctiveness in sentence processing. As stated in the introduction, these models are not intended to replace incremental parsing mechanisms, but rather enhance our understanding of the additional cognitive mechanisms likely to be involved in sentence processing. I will begin by presenting a broad overview of models of memory and the corresponding processing hypotheses that were generated by "decay" based models of memory, which aligns closely with the discussion provided by Van Dyke and Johns (2012). I will then offer evidence in favor of more current interference based models of memory and processing, in particular, cue-based models of retrieval. In continuation, I will highlight some of the gaps in cue-based retrieval theory and discuss alternative sources of memory-related processing difficulty, primarily via encoding interference. I will conclude by presenting the general hypotheses that are addressed in this dissertation.

2.2 Human Memory Architecture

As discussed in Chapter 1, online sentence comprehension must rely on some component of memory in order to hold onto previously encoded linguistic information that might later be used in the resolution of a grammatical dependency. Crucially, these dependencies can be non-adjacent or "long-distance," which implies that the number of items that must be actively maintained in memory may vary from structure to structure. In the cue-based retrieval literature, the term Working Memory is often used to explain the temporary storage space in which linguistic information can be stored and retrieved during online processing procedures. The question then

becomes, what exactly is Working Memory, and what evidence do we have that this system (or a similar system) is employed during online processing?

Classic models propose that the human memory system consists of a tripartite architecture containing Long-Term Memory (LTM), Working Memory (WM), and a small set of representations that fall within the span of focal attention (McElree, 2001). The Working Memory system has typically been isolated from the other two levels of memory due to its capacity to temporarily retain a small number of items that fall outside of the immediate scope of focal attention. Moreover, within WM, items remain more accessible at the point of retrieval than items stored in LTM (McElree, 2001). Several claims have been made to explain the ease of accessibility of the items within WM. Two of the primary claims about increased accessibility are (1) unique storage structures exist for recently activated items (Baddeley, 1986, Baddeley and Hitch, 1974; Schneider and Detweiler, 1988; Shallice and Vallar, 1990; cited in McElree 2001) and (2) recently processed items inherently maintain a higher activation status that makes them easier to access (Anderson, 1983; Conway & Engle, 1994; Engle, 1996; Ericsson & Pennington, 1993 and others reported in McElree, 2011).

Additionally, supporters of the tripartite architecture predict that retrieving items from each level of memory should result in distinct retrieval speeds, as the items are more difficult to retrieve from the deeper levels of memory (McElree, 2001). There is a growing body of evidence, however, that demonstrates that working memory and focal attention are the same system, a few experiments have found time differences in retrieving items from these supposedly distinct levels of memory (McElree, 2001). Regardless of the assumed memory architecture, at some level of human memory, there is a temporally or attentionally constrained space where information can be rapidly processed before being encoded into a deeper level of memory or completely degraded. It is

assumed that a maximum of 1-4 items can be maintained in the span of attention, (four items Cowan, 2001; Cowan, 2016; 1 item McElree and Dosher, 1998; McElree, 2001). This limit could be expanded to include more than one item, under the most restrictive account, if the events can be encoded into one related chunk (McElree and Dosher, 1998; McElree, 2001). For our purposes, the items that are activated within this temporary space (either a distinct WM system or the span of focal attention) are considered to be within the retrieval threshold (Van Dyke and Lewis, 2003), and their behavior within this temporary storage space will be the primary focus of the upcoming discussion.

The experiments conducted in this dissertation are not intended to differentiate between these two views of the memory architecture, so I make no claims or predictions regarding these systems. With that being said, in the upcoming discussion, I will be referring to the Working Memory system rather frequently, as this is the system most commonly discussed in the cue-based retrieval literature. I remain relatively agnostic towards my views on the memory architecture itself and present the following information primarily to express that the human memory system contains a space in which information can be temporarily encoded and accessed within a very short span of time. It is within this space that the retrieval processes related to online processing predicted to occur.

2.3 Decay versus interference as a source of memory failure

Few researchers would deny the claim that some form of Working Memory or related temporary storage system is required for many cognitive processes, including language processing, but the memory mechanisms that underlie successful linguistic retrieval are a source of debate. More specifically, the cause of *forgetting* items that have been stored in memory has yet to be fully understood. Historically, this debate has been divided into two camps: decay and interference.

Following the presentation styles of Van Dyke and Johns (2012) and Parker et al. (2017), the following sections will present an overview of the early evidence that led researchers to believe that the allocation of working memory resources was primarily dependent on storage time (decay) and how this influenced sentence processing theories. Afterward, I will present the growing body of evidence that demonstrates that retrieval failure is more dependent on the content being actively maintained in memory and the degree of overlap between items (interference).

2.3.1 Decay and Working Memory

Early studies on the limitations of working memory claim that the passage of time is the primary cause of *forgetting*, known as the trace-decay hypothesis. The central claim of this hypothesis is that the trace of an item in memory becomes increasingly difficult to access as time passes (Brown, 1958). Essentially, working memory resources cannot maintain early elements in the span of focal attention as new information is being processed. Empirical support for this hypothesis was notably found in experiments by Brown (1958) and Peterson and Peterson (1959). In the Brown-Peterson tasks, participants were presented with a short combination of letters or numbers (e.g. in Peterson would look like *878S-9*) and were expected immediately enunciate the number at the very end. The participants then had to immediately perform an algebraic task, such as counting backward by three from the number that they said out loud. After an experimentally manipulated amount of time, participants were asked to recall the original letters that they read. These early studies found a positive correlation between the passage of time and forgetting, such that items became harder to remember as the time between encoding and retrieval increased, especially if the participants were prohibited from verbally rehearsing the items that they were asked to remember.

The findings in the Brown Peterson experiments, as well as a series of other experiments that claimed that information is rapidly lost from memory unless it is verbally rehearsed, were

explained by limited WM resources (see Van Dyke and Johns, 2012 for discussion). These findings helped to contribute to one of the most referenced models of working memory, the Baddeley (and Hitch) Model of Working Memory (e.g., Baddeley and Hitch 1972, Baddeley, 2002). Although this model has undergone several revisions over the last 50 years, it was initially formulated as a three-part system containing (1) the phonological loop, which deals with acoustic information, (2) the visuospatial sketchpad to process visual information, and (3) an "attentionally-limited control system" known as the central executive (a fourth component known as the episodic buffer has been posited for newer models, Baddeley, 2002). In the Baddeley model, the phonological loop is the most relevant to psycholinguistic research. Baddeley and Hitch's model(s) claimed that information would rapidly decay from short term memory if the information was not verbally rehearsed, thus suggesting that phonological reactivation was a way to prevent information loss (1972). I mention the Baddeley model here, not in support of the model as a description of the working memory architecture, but rather because I find it these factors to be early signs in favor of interference-driven constraints on the memory system (Kush et al., 2015).

In several discussions of the phonological loop and its limitations, Baddeley and colleagues found that increased degrees of acoustic similarity between the items decreased subjects' ability to retain and recall information, even more so than similarity in meaning (Baddeley 1996a, discussion in Baddeley, 2003). This pattern shifted, however, when subjects were asked to memorize lists of 10+ words with related meanings and recall them across trials. In the studies containing 10+ words, the authors found that meaning overlap negatively impacted recall ability more than phonological overlap (Baddeley et al., 1966b, Baddeley, 2003). Important for the upcoming discussion about decay and interference, both of these claims show that linguistic information that is shared between items affects retrieval accuracy. These overlap effects cannot be accounted for by assuming that

time is the only factor in forgetting. It is unclear what role the phonological loop actually plays in models of sentence processing. Still, the data that has been used to support the existence of this mechanism is quite relevant when discussing whether interference or decay is the primary source of memory failure.

2.3.2 Storage and Decay Effects in Sentence Processing Models

Although psychologists at the time made strong arguments in showing how limited working memory storage space and temporal decay effects lead to the degradation of items stored in memory, psycholinguists were faced with the challenge of incorporating grammar into these more general cognitive hypotheses. In response, several memory-related hypotheses were put forth to explain why specific grammatical structures, such as center embedding (3), were more difficult to process than other embedded clauses (2), although they are both permitted by the grammar (examples from Gibson, 2000). In the examples below, the bracketed regions highlight an additional syntactic dependency and the underlined elements represent the matrix subject-verb dependency.

(1) The reporter disliked the editor.

(2) The reporter who [the senator attacked] disliked the editor.

(3) The reporter [who the senator [who John met] attacked] disliked the editor.

Evidence similar to the examples above has generated several hypotheses. Many early hypotheses about the role of memory in sentence processing share the underlying assumption that syntactic structures containing multiple unattached constituents, arbitrarily separated from their attachment site by intervening information, are more difficult to comprehend because we cannot actively maintain unattached constituents in memory for the amount of time that it takes to process the intervening information (Bever, 1970; Gibson, 1991, 1998; Gibson and Thomas, 1999; Kimball,

1973; Lewis 1993, 1996; Miller and Chomsky 1963; see Gibson and Thomas 1999 for a discussion). One of the earliest discussions can be found in Miller and Chomsky (1963). In their article, the authors explained that center embedding, although grammatical, was more difficult to process due to the number of incomplete syntactic dependencies that must be stored in working memory, which then place greater demands on WM resources. This ties into the discussion on decay, although it is not explicitly mentioned, because the strain of processing new information while maintaining the memory traces of the unattached constituents degraded the traces of the early encoded constituents.

Another prominent linguistic theory that employed the concept of limited memory resources and decay is the Dependency Locality Theory (DLT - Gibson, 1998; Gibson, 2000), in which comprehension is driven by storage, retrieval, and integration of information. DLT was constructed partially to explain why certain embedded constructions (4, 5) are easier to processes than (3), despite being structurally similar (3). In particular, Gibson was interested in explaining why having a pronoun as the embedded subject of a relative clause makes sentences like (4, 5) easier to process (Bever, 1970, Gibson, 2000; Warren and Gibson, 1999)

> (4) A book [that some Italian [that I have heard of] wrote] will be published soon by MIT press (Frank, 1992)
>
> (5) The reporter who everyone that I met trusts said the president won't resign yet. (Bever, 1974).

DLT was the first hypothesis to postulate a direct correlation between an information integration cost and structure, claiming that memory-related processing difficulties are not only due to storage costs but also to the grammatical function and integration of the information in the sentence. Gibson proposed a distinction between two types of integration: structural and discourse. Structural integration is defined as the incorporation of an XPs syntactic category into a structure

that has already been built. Structural integration is closely related to other analyses of incremental parsing, such as those discussed in Frazier (1978) and McElree and Griffith (1995), which show that structural attachments "precede contextual plausibility evaluations" (Gibson, 2000).

The second component of the DLT is discourse integration, which is based on findings showing that the "difficulty to processing an NP depends on the accessibility of the referent of the NP in the discourse (Haviland and Clark, 1974, Haliday and Hasan 1976; Garrod and Sanford, 1977, 1982 and Garrod and Sanford 1994 – referenced in Gibson, 2000). Essentially, focused elements, such as proper names and definite descriptions that are often referred to by pronouns, are highly accessible and require fewer resources to process. Additionally, Gibson (1998) claims that discourse processing costs are only present when processing *new* discourse referents. Therefore, a center embedded sentence like (3) is more difficult to process than constructions like (4, 5), in part, because there is an additional and *new* discourse referent in (3) that is not present in the other two structures. This new referent increases the overall integration cost for (3), despite this sentence having similar syntactic integration costs to sentences (4 and 5).

In sum, the previous sections have provided a brief overview of traditional approaches to sentence processing difficulty and the memory mechanisms involved in this process. If early approaches to working memory, which depended on the allocation of limited resources over time, were applied to the grammar, the theories of nesting complexity and the DLT appear to be logical manners of connecting memory to linguistic theory. The following section, however, points out several experimental confounds in early memory studies that have demonstrated that temporal decay and limited memory capacity have minimal empirical support in psycholinguistic research. Moreover, a growing number of psycholinguistic studies have found that a wide variety of

linguistic information appears to complicate the retrieval process, which can't necessarily be explained by integration costs.

2.4 Interference Approaches

As discussed above, the concepts of decay and linear distance were quite influential in the development of general theories of both memory and sentence processing. However, an alternative line of research began to gain momentum in the 1960s, in part due to a confound that was discovered in the Brown-Peterson tasks – a topic has been of particular interest to psycholinguists since the early 2000s. The role of *interference* as a primary source of retrieval failure was proposed when Keppel and Underwood (1962) conducted a modified version of the Brown-Peterson tasks and found that the passage of time did not affect forgetting, so long as no interfering information had been previously presented to the subjects. Interfering information, for the purpose of this discussion, can be defined as distractor items that share a feature (linguistic or otherwise) with the intended target. In both the Keppel and Underwood study and the Brown-Peterson tasks, the "interfering information" was the experimental trigrams that appeared prior to critical trials and were highly similar to the intended target item. Keppel and Underwood found that if the critical trials contained letters that overlapped with those presented during a previous trial, subjects had greater difficulty recalling the critical trial.

In addition, the authors demonstrated that that recall accuracy was higher at the beginning of the experiment since participants inevitably encoded more interfering distractors as time went on. In fact, the first experimental trial was recalled with nearly 100% accuracy regardless of the amount of time spent between presentation and testing, which starkly contrasts with claims that the time between encoding and recall is the sole cause of forgetting. They explained these results by stating that, as the experiment progressed, participants were required to store more highly

related items in memory, which then interfered with the information presented in each subsequent trial. The overlap between the stored items and incoming information decreased the participants' ability to identify the intended target among previously encoded distractors.

The Keppel and Underwood experiments indicated that items stored in memory are in competition with each other, and this competition leads to retrieval difficulty. In the broad context of human memory research, this study demonstrated that competition between items has a greater effect on forgetting than temporal decay. Support for these conclusions has been found in both psychology and psycholinguistics, which encouraged researchers to abandon a purely temporal/decay-based account of WM memory constraints in favor of a model that considers interference as the main source of forgetting (Berman et al., 2009; Obrerauer and Lewandosky, 2013; Lewandowsky et al., 2009; Lewis et al., 2006, and Van Dyke and Johns, 2012 for an overview).

It is important to note, however, that the effects of temporal decay are still included in computational models of sentence processing that focus primarily on interference (e.g. Lewis and Vasishth, 2005). In these models, decay is represented as a constant that decreases the overall activation levels an item simply due to the amount of time that has passed. What is important to note, however, that the current models do not consider decay to be the *only* source of retrieval difficulties. The retrieval literature focuses on the role of interference within "retrieval threshold," which is the time span in which decay effects have not degraded a memory trace beyond retrievability (see Lewis and Vasishth, 2005; Van Dyke and Lewis, 2003; and Lewis et al., 2006 for discussion of computational models). The concept of a retrieval threshold is worked into interference approaches by claiming that longer storage times cause items to experience increased amounts of interference from preceding items. This effect is simply a result of focus shifting away

from early-encoded elements, as opposed to claiming that the trace has been degraded solely due to the reallocation of limited memory resources (Martin and McElree, 2008; Lewis and Van Dyke, 2003). In other words, decay is still considered to be a factor in both interference and storage models, as both predict that long distances between a target and its probe will weaken the retrieval process, but the psychological source of this effect is unique (Martin and McElree, 2008). In this dissertation, I hope to corroborate the claim that similarity-based interference contributes to retrieval difficulties in sentence processing, but I make no claims about the role of decay in working memory as it pertains to linguistic retrieval, as these experiments were not designed to address this question.

2.5 The Cue-Based Retrieval Mechanism

As the domain-general and psycholinguistic evidence in the previous sections demonstrates, there is a large body of evidence that domain-general memory recall is largely constrained by competition between items in memory. However, I have yet to discuss the cognitive mechanisms that have been proposed for retrieving *linguistic* information from within memory. In this section, I will present the cue-based retrieval framework, which presents a theoretical account for both the mechanism(s) responsible for retrieving linguistic items from memory, as well as predicts how and where interference effects will arise during online processing.

As discussed in Chapter 1, the underlying assumption of cue-based retrieval models is that a retrieval cue undergoes a rapid, parallel, associative cue-matching procedure with any and all items in memory that possess a linguistic feature that overlaps with the cue (e.g., Van Dyke and Lewis, 2003). For example, if a verb projects retrieval cues for a plural subject, the item retrieved from memory must contain the proper Case and number features in order to be retrieved as a potential antecedent. I reiterate this concept here, as several of the core developments of this theory

refer to the terms *cue* or a *cue-matching*, so it is important to have the basic framework in mind before presenting the evidence.

2.5.1 Serial versus Parallel Retrieval

As outlined in section 2.3.2, early models of the language-memory interface focus on how linear distance and the number of items intervening between dependent elements decrease retrieval accuracy and speed. A serial-search retrieval mechanism was proposed to explain why distance and the number of items stored in memory were able to decrease retrieval accuracy and speed. The central prediction of serial retrieval models is that retrieval time is correlated with the number of items stored in memory; therefore, as the number of items in memory increases, so does the time it takes to search through them (Sternberg, 1966, 1969, 1975; McElree and Dosher, 1989, 1993). In a sentence processing context, this would imply that it would take longer to resolve the grammatical dependency in (7) than it would in (6). This is because the increased amount of time required to process the intervening relative clause in (7) makes the earlier-encoded items (e.g., the matrix subject) more suspectable to temporal decay by the time the reader arrives at the matrix verb dependency site. The degraded memory trace is more difficult to retrieve, thereby resulting in slower reading and/or reaction times in sentence (7) than in (6).

(6) The students study at the coffee shop.

(7) The students [who were writing their dissertations] study at the coffee shop.

More generally, a serial search procedure implies that each encoded constituent is scanned by the retrieval mechanism during sentence processing. If this were the case, the search process would end once the mechanism has identified a constituent that meets the grammatical requirements of the unresolved dependency (the probe) (McElree, 2000; Martin and McElree, 2008). For example, a verb seeking out a noun with a specific agreement features would scan <u>all</u> encoded constituents

(not just the nouns) until it found an NP that satisfies its linguistic requirements, (e.g. [+Nominative])). Intuitively, the further away the antecedent is from the unresolved dependency site, the longer it will take to comprehend the sentence, as there would be more information to search through. However, a growing number of experiments have used Speed-Accuracy Tradeoff (SAT) paradigms to discredit search-based retrieval models of sentence comprehension.

In SAT experiments, participants are presented with a sentence word-by-word in a Rapid Serial Visual Presentation (RSVP) task and then asked to provide multiple binary grammaticality judgments about the sentence at predetermined intervals. These cued intervals occur throughout the sentence, even at points where the reader does not have enough information to evaluate the sentence's grammaticality. The structure of these tasks is particularly useful in determining when the subject is confident enough in their response to make a decision (e.g. regarding its grammaticality). The cued response intervals occur prior to the onset of the critical dependency and continue for several seconds after its presentation in order to build a comprehensive temporal profile of online processing decisions and interpretations. This method provides empirical data showing the exact point in the sentence where subjects have retrieved sufficient information for the sentence to be deemed as grammatical. As summarized by Parker et al. (2017), the average performance for each participant across the cued intervals is transformed into an exponential curve that reflects the speed-accuracy tradeoff function.

Many SAT experiments focusing on sentence comprehension indicate that an increased number of items in memory only affects response accuracy and not retrieval time, crucially suggesting that retrieval is not serial in nature (see Foraker and McElree, 2011 for an overview of SAT experiments; Lewis and Van Dyke, 2003; Martin and McElree, 2008; McElree, 2006). These results directly challenge the predictions made by decay-based processing theories. Decay-based

processing approaches assert that retrieval difficulties are driven by the linear distance between probes and targets, as determined by surface word order. Essentially, the more intervening information between a target and probe, the more memory resources must be allocated to holding onto the target. If linear distance were indeed the sole cause of retrieval inaccuracies, the retrieval mechanism must be serial, as this is the only way that distance should affect the time it takes to retrieve items. Given the SAT evidence, the predictions made by the early processing hypotheses relying on serial retrieval are largely unsubstantiated.

2.5.2 Content Addressability and Direct Access

Given the strong evidence in favor of a parallel retrieval mechanism, the next step in the development of a retrieval model is to identify *how and which* items are activated and paired with a retrieval cue. To address this matter, psychologists have proposed a content-addressable retrieval model that has direct access to only (and all) relevant memory representations without the need to sort through extraneous information (see Van Dyke and Johns, 2012 for a review of models). Crucially, content-addressable retrieval is assumed to be domain-general and have gained empirical support through linguistic and non-linguistic tasks (e.g., McElree, 2001). Content-addressable models provide further details about how parallel retrieval is operationalized at the psychological level because they offer an explanation of how memory content is activated and paired with a retrieval cue: via the overlap between the content of encoded information and a retrieval cue (McElree, 2000; McElree et al., 2003; Martin and McElree, 2008). Although these models were not originally developed to account for sentence processing phenomena, they possess strong explanatory power in the realm of language comprehension.

To further specify how content addressability is employed in linguistic retrieval, Martin and McElree (2008) claim that content-addressable retrieval allows the comprehender to "consider

only the constituent (in the case of an unambiguous expression) or constituents (in the case of an ambiguous expression) that are fully compatible with all properties needed to resolve the dependency." In other words, the retrieval cue will only match with potential antecedents that meet its requirements based on the content shared between the cue and its target. Moreover, arguments have been made that content addressability is involved in the retrieval of items from both short- and long-term memory (McElree, 2006; McElree and Dosher 1989). This claim is important since the precise memory architecture involved that employs cue-based retrieval is still a source of debate. As pointed out in section 2.1, it is unclear if the memory architecture is tripartite or a combination of long-term memory and attentional activation, but content addressability fits into both memory models without entering into the debate regarding architecture. Essentially, content addressability allows the retrieval cue to have direct access to its target (and potential competitors) regardless of the level of memory that the item is being retrieved from.

Content addressability and parallel retrieval have significant empirical support from SAT paradigms, but researchers must still determine the type of *content* that is involved in linguistic retrieval processes (SAT experiment overview: Foraker and McElree, 2011). The content of an item is the information that is encoded into the memory system, which in sentence processing, is assumed to be a wide range of linguistic features. McElree and Martin (2008) claim that the information available at the site where the dependency that is being processed likely consists of a variety of linguistic content, ranging from morphosyntactic and semantic features to broader semantic and discourse information, based on previous studies that have found a direct link between this information and online processing. This linguistic content can act as a cue, which allows for the retrieval mechanism to directly access all relevant information stored in memory (Martin and McElree, 2008 discussion of McElree, 1996; 1998; 2000; 2006; McElree and Dosher,

1989, 1993). Although cues will be discussed further in the upcoming section, the concept that a content-addressable retrieval model allows retrieval cues to associate with any and all items that contain the desired feature is essential to present early.

2.5.3 Cue-based Retrieval

The overwhelming empirical evidence supporting a content-addressable retrieval mechanism has led to the development of cue-based retrieval models specified for sentence processing (e.g., Van Dyke and Lewis, 2003; Lewis et al., 2005). Cue-based retrieval models provide a theoretical account of the cognitive mechanism that underlies the retrieval of linguistic information from memory, as well as makes clear empirical predictions that can aid in the explanation of several psycholinguistic phenomena associated with the processing of grammatical dependencies . Assuming that content addressability is the primary means by which a cue is able to access its intended target among competitors, cue-based retrieval models can be used to further specify the *mechanism* responsible for the cue-target matching procedure. In the cue-based retrieval framework, it is widely assumed that all potential targets for a given retrieval cue are accessed via a parallel cue matching procedure, in which the retrieval cue has direct access to all relevant items. Although the exact definition of a retrieval cue is still being refined in the literature, in sentence processing, there is evidence that a retrieval cue is composed of the grammatical information, context and structural information associated with a specific word, which all form a subset of the features of the intended target (e.g., Lewis et al., 2006 – revisited in section 2.8). I say this to clarify that I do not assume that cue-based retrieval is *replacing* a sentence parsing mechanism that incrementally builds structure, but rather, that the sentence processing mechanism(s), like many other cognitive systems, are subject to the limitations of human memory such as interference.

In line with the interference accounts of memory constraints (e.g., Keppel and Underwood, 1962), the cue-based retrieval mechanism is sensitive to competition between items in memory. Under this framework, items encoded in memory compete with each other when their content overlaps with the retrieval cue. This competition is driven, in part, by the content addressable nature of this mechanism. In situations where multiple items in memory overlap in content with the retrieval cue, the retrieval cue is said to become "cue-overloaded" (definition: Van Dyke, et al. 2014; concept found in Watkins and Watkins, 1975, 1976). In cases of cue overload, the cue automatically undergoes matching procedures with all items that overlap, even partially, with the content that the cue requires to retrieve its antecedent. The empirical effects of cue overload are known as similarity-based interference, and these effects are particularly useful when testing the interaction between known memory mechanisms and sentence processing phenomena. For example, if a certain type of linguistic content overlap generates similarity-based interference and another does not, claims can be generated about the involvement of specific linguistic content in the retrieval process. Importantly, it is only in cases where there is cue overload that the retrieval mechanism is predicted to hinder sentence comprehension, even though retrieval must always occur to some extent during online processing as items (even adjacent ones) are encoded and retrieved (discussed in Chapter 1).

A concrete example of how cue-based retrieval is employed in language comprehension can be observed in subject-verb dependency resolution (Dillon et al., 2013; Lewis and Vasishth, 2005; Lewis et al., 2006; Wagers et al., 2009 among others). In (8), the matrix verb (*ran away*) projects a retrieval cue for an antecedent with the following features: [+Nominative Case] and [+Animate]). The content-addressable nature of the retrieval mechanism only allows the

[+Animate NP] (*the thief*) to be accessed by the retrieval cue, and the dependency is resolved without any competition from *the cameras*.

(8) The thief that the cameras captured ran away quickly.

$$\begin{bmatrix} +\text{NP} \\ +\text{Animate} \\ +\text{Nominative} \end{bmatrix} \qquad \begin{bmatrix} +\text{NP} \\ -\text{Animate} \\ -\text{Nominative} \end{bmatrix} \qquad \begin{bmatrix} seeking \\ +NP \\ +Animate \\ +Nominative \end{bmatrix}$$

It is not always the case, however, that we encode only one item in memory that overlaps with the retrieval cue. As Van Dyke and colleagues (2011, 2012) point out, the origins of interference are rooted in general human memory research, but a rapidly growing body of research shows that interference effects disrupt sentence comprehension (see Van Dyke and Johns 2012 for an overview of experiments). For example, in (9), two animate NPs have been encoded, one which serves as the subject of the relative clause (*the thief*) and one that is the intended subject of the matrix clause (*the officers*). Although both sentences are structurally identical, in (9), the retrieval cue may be unable to unequivocally and/or quickly match with its antecedent, as there is cue overlap between the two [+Animate] NPs, thus generating competition between these items. This competition between NPs can result in similarity-based interference effects, causing sentence (9) to be harder to process than (8):

(9) The thief that the officers captured ran away quickly.

$$\begin{bmatrix} +\text{NP} \\ +\text{Animate} \\ +\text{Nominative} \end{bmatrix} \qquad \begin{bmatrix} +\text{NP} \\ +\text{Animate} \\ +\text{Nominative} \end{bmatrix} \qquad \begin{bmatrix} seeking\ NP \\ +\text{Animate} \\ +\text{Nominative} \end{bmatrix}$$

Although these examples are simplified, they are intended to show how retrieval interference might occur, assuming a cue-based retrieval mechanism.

2.5.3.1 Models of cue-based retrieval

With the cue-based retrieval framework presented, I will now focus on the model(s) of cue-based retrieval that I will be assuming throughout this dissertation. Although I will not commit myself to a single model (e.g. Lewis and Vasishth, 2005 vs. Lewis and Van Dyke, 2003), as no single model can account for all of the open questions related to retrieval, I want to present the baseline theoretical assumptions that I will reference throughout the thesis. To begin, I assume the baseline retrieval procedure presented in the previous section, where a retrieval cue is directly associated with a target that is activated in memory based on the content that overlaps between the two. In addition, when multiple items in memory contain the same content, the retrieval cue becomes cue overloaded and similarity-based interference effects will occur, resulting in empirically observable processing delays and inaccuracies (e.g. Lewis and Vasishth, 2005; Martin and McElree, 2008; Parker et al., 2017, 2019l Van Dyke and Lewis, 2003; Van Dyke and McElree, 2006; Van Dyke and Johns, 2012, among many others). Moreover, for the sake of this dissertation, I will be focusing on morphosyntactic features, likely a subset of the linguistic content that is employed in the retrieval process, to test for and describe interference effects in Brazilian Portuguese sluices and Relative Clauses. I acknowledge that the current cue-based retrieval models are largely incomplete (Parker et al., 2017) and present the assumptions above without committing to a specific model. The purpose of this dissertation is to employ the standard assumptions of cue-based retrieval to a novel discussion regarding *feature relevancy* in hopes that the findings help to arbitrate between current proposals.

In an attempt to identify the components of the retrieval mechanism, Lewis and colleagues (2006) stated that the most effective model requires five, if not more, key computational principles to function within our current understanding of the memory architecture:

i. Extremely limited focus of attention;
ii. Fast content-addressed access to item information but not serial order information;
iii. Similarity-based retrieval interference;
iv. Fluctuating activation as a function of decay and retrieval history;
v. Similarity-based encoding interference.

As the authors point out, principles i-iv have been mathematically implemented in the adaptive control of rational thought (ACT-R) model (Anderson, 2004; adaptation by Lewis and Vasishth, 2005). More recent adaptations of the ACT-R model have been generated to account for the fifth principle, encoding interference, and this dissertation also provides some empirical evidence that encourages the inclusion of this principle into standard models (e.g., Villata et al., 2018). Although several computational models of cue-based retrieval have been proposed, I will frequently refer to the adaptation of ACT-R produced by Lewis and Vasishth throughout this dissertation, not only due to its success at replicating and predicting real-time processing patterns but also due to the ability to adapt the model to factor in encoding interference effects and its consideration of a parsing mechanism (e.g. Villata et al., 2018). I acknowledge that the original ACT-R model was not intended to describe language-specific phenomena and that its series of buffers do not necessarily align with current models of memory that do not contain discrete storage systems (see discussion of memory architecture in section 2.2).

ACT-R is a highly developed computational models of human rational thought, and recent adaptations to the model have developed a method of activation distribution that serves as a proxy for similarity-based interference (Anderson et al., 2004; Anderson, 2005 for updated discussion; Lewis and Vasishth, 2005 for discussion on sentence processing research). Focusing purely on the sentence processing adaptation, this framework contains both an incremental ("left-corner) parser and allows targets to be retrieved via their activation levels. The activation level of a chunk is

traditionally attributed to three main factors: (1) the base activation level of the chunk, which is influenced by the chunk's past activations and the number of times that it has been re-activated, (2) the strength of the association between the retrieval cue and the chunk, which accounts for the uniqueness of the chunk in comparison to potential competitions during the retrieval process and (3) random noise . From this point, the model relies on a winner-take-all approach where only the chunk with the highest activation level is eventually retrieved.

Assuming that retrieval is a winner-take-all system where there is only *one* correct target (e.g. only one subject for a given verb), the ACT-R model treats the item with the highest level of activation in memory as the most likely target for the dependency in question. This means that no matter what, the system will only select a single target, though the "race" to select the target is affected by competitors/distractors with similar activation to the intended target. This pairing slowdown is intended to replicate online measures that show that reading times decrease when interfering competitors have been stored in memory. Crucially, this does not mean that the sentence is ultimately misparsed online or that the retrieval mechanism is overriding the grammar, but rather demonstrates that competition between items decreases the efficiency of the retrieval process and overall online processing speed. In extreme cases, this competition appears to lead to a sentential misparse, where the incorrect target was selected, but this does not necessarily mean that the sentence was parsed incorrectly online (Schlueter et al., 2019). There are a variety of post-sentential processes that can contribute to an offline interpretation being different from the online parse, and the experiments in this dissertation do not use sensitive enough methods or designs to disentangle whether misretrieval or post-sentential processing led to some of the offline results presented in Chapters 3 and 4. More commonly, however, the correct interpretation of the sentence is obtained, but reading and reaction time measures and comprehension question accuracy are

decreased (variety of examples in Van Dyke and Johns, 2012). These effects are examples of similarity-based interference.

In order to describe how more than one chunk can possess similar activation levels, the ACT-R model contains a term to account for the *fan effect*. Under Anderson's approach, the fan effect refers to the phenomenon that retrieval time (or processing time) increases as subjects learn more about the given topic (Anderson 1974). In the linguistically focused ACT-R model, this fan effect is modified to account for retrieval interference: "Associative retrieval interference arises because the strength of association from a cue is reduced as a function of the number of items associated with the cue" (Lewis and Vasishth, 2005). Essentially, the activation of the linguistic feature that matches the retrieval cue is spread across competitors, making the association between the cue and the intended target much weaker. This weakened associative process causes the selection of the target to be slower and/or less accurate. Although the research presented in this dissertation cannot support the accuracy of ACT-R modeling with respect to human cognition, these models have been highly successful at representing reading time data and sentence processing biases. Moreover, the language comprehension adaptation of ACT-R contains a left-corner parser, which is a computationally defined incremental parser that combines properties of both a bottom up and top down parsing routines. The inclusion of a parsing mechanism in these models highlights the likelihood that sentence comprehension involves both a structure building and a memory mechanism, although this dissertation is not intended to determine the exact points of crossover between these mechanisms. For these reasons, ACT-R adaptations appear to be a fruitful way to model how cue-based retrieval and interference can be applied to sentence processing.

2.6 Types of Similarity-Based Interference

As discussed, similarity-based interference during retrieval occurs when the retrieval cue is unable to immediately and/or unambiguously identify its intended antecedent among competitors stored in memory. Although some of the most well-known cases of retrieval interference are similar to example (11), such that the distractor intervenes between the target and the retrieval cue, several studies have shown that interference can occur with different target/distractor orders as well. In the literature, similarity-based interference can be further categorized as being either *proactive* or *retroactive*. Proactive interference occurs when the intended target is placed between a distractor and the retrieval probe (Distractor – Target – Probe), whereas *retroactive interference* occurs when the distractor is placed between the target and the probe (Target – Distractor – Probe). The image below, from Van Dyke and Johns (2012), provides an illustration of the two types of retrieval interference, where one could assume that the brackets contain the linguistic features that have been encoded for each word.

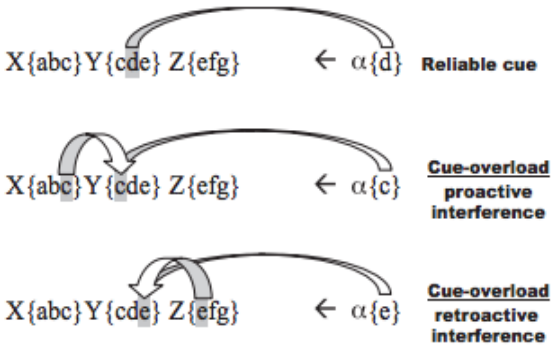


Figure 1: *Types of Similarity-Based Interference (Van Dyke and Johns, 2012)*

In sentence processing, both retroactive and proactive retrieval interference effects have been observed (*Retroactive*: Harris, 2015, 2019; McElree et al., 2003; McElree, 2009, Van Dyke and McElree, 2011, Sasson, 1971, etc.; *Proactive*: Fedorenko et al., 2006; Gordon, 2002; Van Dyke and McElree, 2006). A study by Van Dyke and McElree (2011) tested both types of

interference to see if one had a greater effect on retrieval failure than the other, as previous studies had stated that retroactive interference effects modulated sentence comprehension more than proactive interference (McElree et al., 2003). In two SAT tasks, the authors found no difference between the types of interference on retrieval speed but rather a difference in how proactive and retroactive interference modulated retrieval accuracy. In cases of retroactive interference, overall sentence comprehension is lower. These findings support the previous claims that processing intervening distractors is more disruptive to the retrieval process than retrieving information that was presented prior to the intended target. With this evidence, the following chapters focus primarily on retroactive interference effects. Although each chapter provides a more detailed explanation of the predictions for where and why retroactive interference effects appear, they share the underlying assumption that sentences containing a distractor between the probe and its target will be harder to process than those that have a distractor preceding the target.

2.7 Encoding Interference

Until now, similarity-based interference has only been discussed as a result of the retrieval mechanism. Although the findings are not consistent across studies, recent experiments have found evidence for another type of similarity-based interference known as *encoding* interference (e.g. Gordon, 2001; Fedorenko et al., 2006; see Villata et al., 2018 and Jäger et al., 2015 for an overview of the literature). Much like the retrieval interference effects discussed in the previous section, encoding interference arises when features overlap between items while they are being stored in memory. Although both types of similarity-based interference result in lower sentence comprehension and slower reading times, the underlying source of interference is theoretically distinct. On the one hand, retrieval interference is expected to occur only at the *retrieval site* – the exact point where the retrieval cue is forced to match with an item in memory to resolve a

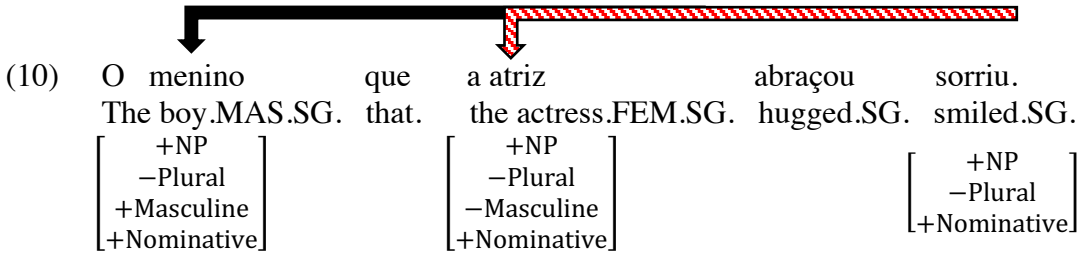
dependency. Encoding interference, on the other hand, is expected to occur *prior* to the retrieval site, merely as a byproduct of attempting to encode items with similar features.

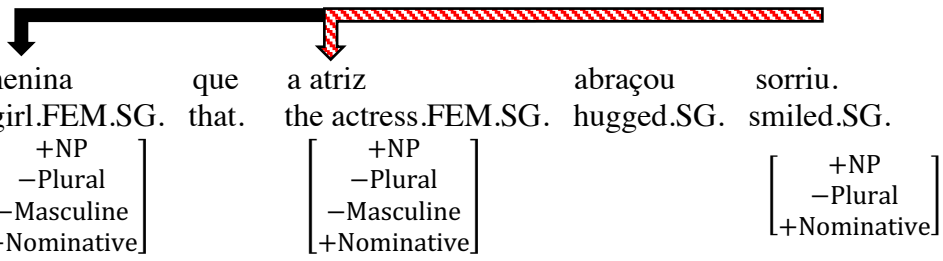
Despite being theoretically distinct, encoding and retrieval interference are quite challenging to empirically disentangle, as they are not always mutually exclusive. This is because features that undergo encoding interference and are thereby degraded in memory, may also be activated during retrieval. As one could imagine, the linguistic features that are used during the retrieval process form a subset of all of the potential linguistic features of a word that have been encoded into memory. With this in mind, it is a logical possibility that some of the effects that are observed at the point of retrieval have *also* undergone encoding interference. For example, two competitors might have been encoded as containing the same gender and number features, which both generate some degree of interference during the encoding stage for each respective feature. This encoding interference could potentially cause some sort of processing disruption before the comprehender even arrives at a retrieval site, which may be a dependency that only requires number information. This crucially means that by the point that the retrieval site is encountered, the number feature has undergone *both* types of interference.

This potential interaction between encoding and retrieval interference make their respective similarity-based interference effects nearly impossible to disentangle in most experimental designs. Essentially, researchers cannot easily determine whether retrieval interference effects have previously been affected by encoding interference or if encoding interference has even occurred at all. Fortunately, there is a relatively simple way to address the latter. Experimenters can test for encoding effects by examining features that are not expected to be involved in the retrieval process (henceforth, *irrelevant or non-retrieval features*). Ideally, the encoding effects should be observed *prior* to the retrieval site, but recent studies have claimed that

the online effects of encoding interference are challenging observe online and are oftentimes conflated with retrieval interference at the retrieval site itself. Crucially, even if the effects are observed only at the retrieval site, if a processing delay occurs due to the match of non-retrieval features, there is evidence that a secondary mechanism is the root of this effect.

To concretize the difference between retrieval and non-retrieval features, I have provided an example of the Portuguese subject-verb dependencies that will be discussed in the following chapter. In Portuguese, verbs are overtly marked for number, and nouns are marked for both gender and number. Therefore, in Portuguese, the feature that is relevant for the retrieval process should be the number feature that is projected by the verb and grammatically agrees with the subject. The gender feature is not relevant to the dependency and should thereby be ignored by the retrieval mechanism under classic models of content-addressable cue-based retrieval. Standard cue-based retrieval models would predict no difference between sentences (10) and (11), even though the gender feature overlaps between the matrix and RC NPs in (11), as the only retrieval features at play should be number and perhaps Case of the NP in question. Although this design cannot answer the question from the previous paragraph about the extent to which number has also undergone encoding interference, it provides a clear-cut way to observe any potential effects of similarity-based interference caused by non-retrieval features.





Along these lines, recent studies on Italian and French relative clauses, as well as several studies on German, Swedish and Russian anaphor resolution, have found that gender matches in environments similar to the one above (discussed in more detail in the following paragraphs) do indeed result in processing difficulties. Although the literature is relatively limited and the results vary across studies, the presence of any sort of processing disruption caused by these irrelevant features merits further investigation. In this section, I discuss the literature on encoding interference and its interaction with the literature on cue-based retrieval, as well as present a current ACT-R model that addresses these concepts (Villata et al. 2018).

Although there is some evidence that similarity-based interference may occur from more than one process within memory (encoding *and* retrieval), the mechanism responsible for encoding interference is still under discussion. One of the most well-known hypotheses about the mechanism behind encoding interference is known as *feature overwriting* (Nairne, 1990). Under this account, interference is primarily driven by features and memory constraints that do not allow for a single feature to be activated on more than one item in memory. A simple way to think about this is that items are encoded into memory with a value of [1]; however, when a new item is encoded that shares this feature with a previous item, the initial item's feature is degraded to [0] due to the inability to maintain two fully marked features on competing items in memory. The trace of the unvalued feature is then degraded in memory, which may negatively impact its retrieval and/or processing time (Nairne, 1990; Oberauer and Kliegel, 2006; Oberauer and Lange 2008; Villata et

al., 2018). On a very rudimentary level, one could imagine that this process behaves like the example below. It is important to keep in mind that the example is meant to show what one might encounter while reading. The leftmost box represented the information before encoding, and the rightmost box shows what is actually encoded into memory with feature overwrite. It is also important to remember that feature overwrite precedes retrieval, which is why the NPs are separated from the verbs.

What the comprehender encounters:	**What the memory system encodes:**
The baby who the mother... loves cries a lot.	<u>The baby</u> who the mother... loves cries a lot.
$\begin{bmatrix}+\text{Singular} \\ +\text{NP}\end{bmatrix}$ $\begin{bmatrix}+\text{Singular} \\ +\text{NP}\end{bmatrix}$	$\begin{bmatrix}\mathbf{-Singular} \\ +\text{NP}\end{bmatrix}$ $\begin{bmatrix}+\text{Singular} \\ +\text{NP}\end{bmatrix}$

Table 1: *Example of Feature Overwrite*

Unfortunately, Nairne (1990) fails to fully spell out the mechanism responsible feature overwriting to a satisfactory level and acknowledges this in the original article. Recent research, however, has incorporated encoding interference via feature overwriting to a variety of linguistic and non-linguistic models of memory with high success rates. One of the most referenced models of feature overwrite can be found in Oberauer and Kliegl (2006), who generated a content-addressable memory model that incorporated feature overwrite prior to retrieval, and also found support for the overwriting mechanism in the synchronous neuron firing models (also mentioned in Jäger et al., 2017). Although this scope of this dissertation does not consider the neurological underpinnings of memory mechanisms, the success of incorporating overwriting into a model of content-addressable memory motivates the upcoming discussion about how an overwriting mechanism could generate to encoding interference effects in sentence process. Findings of this type, showing that feature overwrite can be a viable explanation for interference effects shown to play a role in sentence processing have also been included in a variety of models focused on language comprehension (e.g.; Vasishth et al., 2017; Villata et al., 2018). In fact, a recent study by

Vasishth and colleagues found that feature overwrite models provided a statistically better fit for interference effects than a purely cue-based retrieval approach, albeit the feature overwrite must be accounted for as a finite-mixture process (see discussion in Vasishth et al., 2017).

Although only a few models that provide concrete claims about how the feature overwriting mechanism operates or its psychoneurological basis, the majority that consider an overwriting mechanism claim that this process directly affects the activation level of an item stored in memory. In the Oberauer and Kliegl (2006) model, a feature that has been overwritten becomes less active in memory or "degraded." Although this model does not directly address retrieval, a clear linking hypothesis can now be made between feature overwriting/encoding interference and retrieval latencies. Assuming that items with higher activation levels are privileged during retrieval (e.g., Lewis and Vasishth, 2005), the lower activation level caused by feature overwriting decreases the retrieval probability associated with the item. This lower retrieval probability, caused in part by encoding interference, exacerbates the processing delays commonly associated with similarity-based interference during retrieval (further discussion in Jäger et al., 2015, 2017). These claims have inspired other further discussions about why unmarked features are harder to retrieve, albeit without directly referring to activation levels and overwriting, but it is not hard to see how feature overwriting could contribute to the markedness of a given item (e.g., Villata et al., 2018). The example below outlines how this might work, abstractly assuming that whichever element is closest to the value of 1 is the one that is most likely to be retrieved.

(12) No encoding interference. Both NPs are encoded at 100% activation.

Sentence: ....NP1 [Pl.] ..... NP2 [Sg.]

Activation Level: .... [100%]. ....[100%]

(13) Feature overwriting has lowered the activation level to 0% for NP1 and 100% activation for NP2.

Sentence: ....NP1 [Pl.] ..... NP2 [Pl.]

Activation Level: .... [0%]. ....[100%]

Knowing that a complete feature overwrite (decreasing activation levels to 0%) could potentially degrade a memory trace beyond the threshold of accessibility, it seems highly unlikely that the encoding mechanism behaves in this manner. One could assume many logical possibilities for how activation levels are distributed among competitors during encoding. The schematization below outlines some of the distinct possibilities for how activation levels are redistributed as a result of *partial* feature overwriting. In case (15a) the activation is leveled across both NPs, in (15b) NP2 is more active, and in (15c) NP1 is more active, but crucially, neither NP has full (100%) activation. Crucially, although I am representing the entire NP's activation level (NP1 vs. NP2), the changes in activation levels are driven by features alone, which then affect the encoding of the entire NP

(14) If NP1 ≠ NP2 → NP1 = 100% and NP2 = 100%

(15) If NP1 = NP2 → ....

(a) NP1 = NP2 < 100%

(b) NP1 < NP2 < 100%

(c) NP2 < NP1 < 100%

To further investigate how an encoding mechanism might be implemented into sentence processing and which of the potential activation distributions is most likely, Villata et al. (2018)

attempted to capture the feature overwriting mechanism in an ACT-R model (sentence comprehension; Lewis and Vasishth, 2005). As previously discussed, the ACT-R model is able to predict a *match* effect for retrieval cues and potential targets/distractors that mirrors the reading and reaction time slowdowns observed in online studies that examine interference effects. In addition, chunks that overlap in features are subject to the *fan effect,* where feature matches cause the activation level of a chunk to spread to other chunks also containing this feature. Through activation spreading, the race to retrieving the correct chunk is slowed down, as it becomes harder to find select the chunk with the highest activation levels. Essentially, it becomes more challenging to retrieve a target when the intended target and distractor have similar activation levels (e.g., 65% for the target and 67% for the distractor, post-spreading) than when the target is significantly more activated than the distractor (e.g., 90% activation for target and 10% activation for distractor). Feature spreading is used to explain the match effects that we observe in online studies, such as slower reading times or lower sentence comprehension levels.

Although feature spreading and the fan effect explain retrieval difficulty, they do not inherently account for encoding interference. As it is currently formulated, the ACT-R model only considers feature spreading on features that overlap with the retrieval cue and make no predictions about the effects of other feature matches. Villata et al. added a fourth component that they refer to as the *leveling effect* to account for the encoding interference. Like the fan effect, the leveling effect causes activation levels to spread across all chunks that share a feature. When a new chunk is encoded that shares a feature with a chunk that was previously encoded, the activation level of the crucial feature is equalized in both chunks. This means that an element that was highly activated will become weaker and vice versa. This is the case of (2a) in the example above. Importantly, the leveling effect will occur on both retrieval and non-retrieval cues and chunks,

which essentially generates some like that fan effect, which only affects retrieval cues, that affect all chunks with feature overlap.

The authors claim that, although feature leveling affects both retrieval and non-retrieval features, a model that contains only feature leveling (an encoding-only approach) does not align with effects showing that retrieval cues increase the strength of similarity-based interference (e.g., their own study on English in the same paper; Belletti et al., 2012; Van Dyke and McElree, 2006). When both fan and leveling effects were included in the ACT-R model, the authors were able to replicate all of the findings from the online reading experiments that they conducted, while acknowledging that it would be more theoretically parsimonious to derive encoding effects and predictions for independent reasons. [3] Another potential flaw with adopting an ACT-R model containing both fan and leveling effects, as discussed in the models of cue-based retrieval, is that ACT-R models do not necessarily align with current memory models that do not contain memory buffers. However, the success of this model at replicating the effects that are predicted by both retrieval and encoding mechanisms with respect to both types of interference provides a convenient way to incorporate both mechanisms into a coherent theory. I acknowledge that, mathematically, incorporating both a level and a fan effect generates "double" interference effects in the ACT-R model, but if both of these processes do indeed occur in sequence, this appears to be one logical possibility of how encoding interference contributes to retrieval interference effects for independent reasons. Further research is needed to see if encoding and retrieval mechanisms can be independently motivated. For now, it appears that the Villata et al. adaptation of the ACT-R

[3] Villata et al. (2018) and Franck et al. (2020) also propose that a Self-Organized Sentence Processing (SOSP) approach could account for both encoding and retrieval effects without the need to refer to ACT-R. Although this approach is of theoretical value, I am focusing on the ACT-R account to align with previous (and well-supported) models of cue-based retrieval. Moreover, SOSP models present challenges for current models of parsing that fall beyond the scope of this dissertation.

model provides a strong "first step" into creating testable hypotheses about when and where we should expect to find both independent and related retrieval and encoding effects. Abstracting away from the specific formulation of a model, the existence of both an encoding and retrieval mechanism could be beneficial when trying to explain interference effects for features that are not expected to influence retrieval, but I do recognize that this would be a less parsimonious theory than simply making adaptations to the retrieval model to account for effects of non-retrieval features.

With a potential theoretical mechanism in mind, I will now transition into *why* encoding effects are so hard to identify. One of the biggest challenges associated with encoding interference is that the effects oftentimes appear on the retrieval site, making the exact source of interference difficult to identify (Jäger et al., 2015; Villata et al., 2018). Often, the findings that indicate encoding interference on the retrieval site are presented as either evidence for encoding interference or as proactive interference. With that being said, the experiments that were designed to specifically address encoding interference are scarce and present a mixed bag of findings. For example, two studies on Italian relative clauses have provided contradictory results about the effect of gender match in online measures when gender was not a retrieval cue (see Villata et al. 2018 for support for encoding in Italian and Belletti et al. 2012 for an argument against encoding).

Aside from varying results within languages, encoding interference has not been found for all types of grammatical dependencies. A notable set of eye-tracking experiments on Swedish and German reflexives, where gender is a non-retrieval feature, failed to find online interference effects. The lack of encoding effects in German reflexives and pronouns were replicated in a self-paced reading follow-up; however, Russian readers showed a different pattern (Laurinavichyute et al., 2017). Russian speakers showed sensitivity to gender marking in both pronouns and

reflexives, as well as subject-verb dependencies, neither of which can be explained by retrieval interference alone (Laurinavichyute et al., 2017). The authors state that the difference between encoding effects found in these languages is likely due to the complexity of the structure and the skill of the readers, as the reflexive/pronoun study in German is more syntactically complex than what was presented to the Russian participants. Additionally, they found that in both studies, the offline comprehension question results show sensitivity to gender interference in German, Swedish, and Russian. These offline findings suggest that regardless of the language, encoding interference effects occur at some point in processing and/or sentence wrap up, even if they do not appear immediately in online measures.

The observation that encoding interference appears more consistently in offline than in online experiments has been echoed in other studies. In fact, Villata et al. (2018) claim that offline encoding effects have been found in nearly all of the studies that found no online encoding interference:

> "All in all, our finding that similarity in agreement features affects the comprehension of grammatical sentences in Italian and English aligns with other adult studies – Franck et al. (2015) and Villata and Franck (2016) in French, Jäger et al. (2015) in German and Swedish – as well as developmental studies in Italian (Adani et al., 2010), English (Adani, 2012), Hebrew (Belletti et al., 2012) and German (Adani, 2012), which all provide evidence for encoding interference effects. Similarity-based interference in all these studies manifested in off-line measures of comprehension accuracy (with the exception of Franck et al., 2015 who found an effect on-line)"

Although online and offline processing are not identical, the consistency of encoding interference effects in offline studies, when coupled with the few studies showing weak online effects, make this topic worthy of further investigation. In the broader realm of sentence processing, having evidence that encoding affects linguistic information will help us make more informed processing models with increased explanatory power.

With all of this being said, I will address the concepts of encoding and retrieval interference throughout this thesis in hopes of contributing to the theoretical formation of both models. From the encoding standpoint, Romance languages present an exciting way to test for encoding interference, as these languages have both gender and number features that undergo distinct agreement procedures, making both features relevant or making one feature (gender) irrelevant for retrieval in different constructions. As we will see in the following chapter on Portuguese Relative clauses, as well similar studies on Italian and French (Villata et al., 2016, 2018), gender matches between matrix and RC nouns generate a slowdown in reading times in Romance relative clauses.[4] This suggests that similarity-based interference effects arise from features that are not relevant for the retrieval process and opens the door for further discussions on encoding.

2.8 What is a cue? What is a feature?

Throughout this review of the literature, I have used the term retrieval cue and feature quite frequently to differentiate between linguistic information that is employed in retrieval and information that simply contributes to our representation of an item in memory. It is important now to discuss what is known in the literature about cues and features before proceeding to a general discussion of the hypotheses that motivated the experiments in the following chapters. This distinction, in fact, has been noted as one of the three biggest issues in cue-based retrieval theory (Parker et al., 2017). To begin to address this question, I will define the terms feature and cue, then describe the findings presented in Parker et al. (2017) and describe current work being done to determine what linguistic information actually constitutes a retrieval cue.

[4] A related study on Italian relative clauses found that gender did not affect RC processing, but I take the cross-linguistic evidence provided by Villata et al. (2018) and this present dissertation as sufficient to raise questions about encoding (see Belletti et al. 2012 for a comparison of gender behavior in Hebrew and Italian relative clauses).

Prior to delving into a discussion of retrieval cues and their behaviors, the distinction between linguistic *features* and retrieval *cues* must be made explicit, as these terms will be referred to frequently throughout this thesis. To begin, a feature is a specific component of the linguistic information associated with a word in a sentence (e.g. number morphology on a verb). Linguistic features are derived from a variety of components of the grammar, such as morphology (gender, number), syntax (c-command, Case), or semantics (noun class type, animacy). In order to avoid direct involvement with theoretical debates on how these features arise, I assume that a word has been assigned all of its linguistic features prior to the point of encoding the word into memory by a grammatical parser. Therefore, the features encoded with the DP *the girls* in (16) would be along the lines of [+Nominative], [+DP], [+Definite], [+Animate], [+Plural], and [+Feminine], although not all of these features are overtly expressed on the constituent itself.

(16) The girls study frequently.

Given the wide range of linguistic information that could, then, potentially be involved in retrieval, I intentionally limit the scope of this dissertation to consider only two morphosyntactic features that are overtly expressed in Brazilian Portuguese: gender and number. This subset of overt features does not preclude the existence of covert syntactic features, nor question the involvement of covert features in retrieval and grammatical agreement procedures, but rather allows for very clear and observable predictions to be made. Unlike English, number is overtly marked on all verbs in Portuguese, which abstracts away from detailed speculation about the underlying grammatical agreement procedures that must occur to associate a subject with a verb that is not overtly marked for number (e.g. English: *The graduate student(s) studied*). It is important to emphasize, however, that the grammatical agreement between unmarked verbs and subjects in languages like English

implies that some sort of grammatical parser, which forms basic structure and identifies the syntactic interpretation of constituents, cannot be replaced by a retrieval mechanism.

With this definition of features in place, I return to the term "cue." The retrieval cue is a projection at the site of an unresolved dependency that signals to the processor that a target must be retrieved from memory. These cues, from a linguistic standpoint, are matched with relevant linguistic features during retrieval to retrieve relevant items (the target) from memory. For example, in (20), the potential retrieval cues projected by the verb *study* require a target that is [+DP], [+Animate], [+Nominative], and [+Plural]. When looking only at the potential target in the sentence, it becomes clear that not all linguistic features presented on the target are related to the retrieval cue itself. Only four of the six features encoded with *the girl* were relevant to the retrieval cue. These four elements, by the definition employed in this dissertation, are *relevant features,* and the remaining elements are *irrelevant features*.

Parker et al.'s (2017) point out that the exact mapping between linguistic features and retrieval cues is a cause for debate. Since this topic is still relatively open, I will outline Parker et al.'s (2017) claims about feature-to-cue mapping and follow up by discussing their claims on cue-weighting. To begin, the authors compare two potential theoretical models for feature-to-cue mapping. The first model assumes that all features expressed on a dependency can be cues and are treated equally at retrieval, which has been adapted from the literature on general memory recognition (Clark and Gronlund, 1996, discussion also in Martin and McElree, 2008). This model assumes that any type of cue overlap (from any linguistic feature) should and will generate interference effects. However, a growing number of studies have shown that interference effects do not appear consistently with certain cues or syntactic constructions, which makes it unlikely that *all* features can become retrieval cues in every context or that these features are unweighted

(review in Parker et al. 2017; Parker, 2019). These findings have led researchers like Parker et al. (2017) and Parker (2019) to begin to explore the topic of cue combinatorics.

To begin, several studies have failed to show predicted interference effects for certain morphosyntactic features and structures when the distractor was not syntactically accessible. For example, Van Dyke and McElree (2011) noted that a verb seeking an animate target only showed interference effects from a distractor item when it appeared in subject position, thereby suggesting that nonstructural features such as animacy are not treated in the same way as structural cues (e.g., subjecthood) during retrieval. In this study, the authors consider the lack of semantic interference when syntactic cues were not also shared to be evidence for a weighted cue system in which syntactic cues receive a higher weight/activation than non-syntactic cues and potentially restrict the set of potential targets. In cases where only weakly weighted cues were matching, interference effects failed to surface. This can be taken as evidence against a sort of additivity approach where the presence of multiple overlapping cues, regardless of their linguistic function, produce stronger interference effects than when only one cue is present.

In contrast with Van Dyke and McElree (2011), another set of studies showed that anaphors do not experience interference effects from morphosyntactic features, which have generated interference effects for a variety of other dependencies (e.g. Dillon et al., 2013; overview in Dillon, 2014; Jäger et al, 2015). If morphosyntactic and syntactic information is *always* heavily weighted and delimits the set of potential antecedents, it is unclear why interference from these features does not surface in anaphors. However, a series of related studies have shown that these effects, or lack thereof, are quite inconsistent. Another study on anaphor processing challenged the findings by Dillon et al. (2013) by showing that reflexives are indeed subject to interference effects when the reflexive and the licensor match in more than one feature (Parker and Phillips, 2016). In essence,

certain non-syntactic cues might not be strong enough to cause retrieval interference on their own but generate interference when combined with other overlapping linguistic features. Parker et al. (2017) point out that similar effects have also been observed for thematic binding and for subject-verb dependencies where the strength of the interference is largely dependent on the structural information associated with the target (e.g. Parker et al., 2015). When combined, these studies suggest that the retrieval mechanism does not treat all linguistic cues in the same way, but rather, assigns weights to certain features depending on their syntactic role. Crucially, these combined studies all share the conclusion that cues are weighted and that syntactic cues receive the highest weight. In addition, weakly weighted cues must be combined with a strong cue for effects to surface.

In a computational approach to cue combinatorics, Parker (2019) claims that the retrieval mechanism may assign weights to certain cues based on their syntactic function, degree of overlap, and diagnosticity. Although the combinatorics models presented in Parker's (2019) article focus primarily on the topic of linear versus nonlinear combinatoric patterns and anaphor resolution, some of the discussion points are highly relevant to the dependencies that I will deal with in this thesis. Stepping away from the mathematical formulation of cue weighting, Parker appeals to the concept of cue diagnosticity and reliability as a means of determining a cue's weight (Martin, 2016; Harris, 2015, 2019).

In simple terms, cue diagnosticity essentially refers to the extent to which a cue, or set of cues, can sufficiently identify the target from its competitors. Various linguistic factors can contribute to the diagnosticity of a cue, such as its discourse and structural function (Harris, 2015, 2019), and under some accounts the probabilistic reliability of this cue in a given language (Martin, 2016). In most discussions of cue diagnosticity, highly diagnostic cues tend to show less

interference that weakly diagnostic cues (Harris, 2015, 2019). Therefore, in cue combinatorics, the ideal combinatoric pattern is the one that makes the set of cues as diagnostic as possible for the given target/probe dependency (Parker 2019). This concept is very important because this suggests that the retrieval mechanism does *not* operate in the same fashion for all dependencies, and we cannot expect the same type of linguistic information to always generate identical interference effects in all structures (Parker, 2019). Moreover, the concept of diagnosticity is one that can be used to estimate the weight of a given cue before testing it in the language at hand. For example, the gender cues that I address in this dissertation are likely to be more diagnostic in Portuguese than in English, due to their grammatical necessity and high probability of use in Portuguese. Keeping these concepts in mind can help generate insightful and testable hypotheses that contribute to successful experimental design and theoretical claims.

The concept of combinatorics is important to explore, as one could imagine that a wide variety of linguistic "features" (morphosyntactic information, syntactic features, and phonological features) could potentially be involved in the retrieval process and weighted in distinct ways. Being aware of how cues are weighted and the roles that these weights play in the retrieval process can potentially help to delimit the set of potential antecedents that are available to the retrieval mechanism. I will return to this discussion in more detail in Chapter 4, which discusses the role of constraints (syntactic or otherwise) on the retrieval mechanism. For now, I present this discussion primarily to show that there is sufficient evidence that not all linguistic information is a strong cue in retrieval, and that the strength of this information is heavily affected by retrieval-external factors.

Although this dissertation will not address the mathematical nature of cue-combinatorics, I will refer to the discussion of cue weighting as I attempt to unravel the difference between a retrieval mechanism with where all cues are relevant, even those that are not overtly expressed, and one that considers only on relevant cues and is influenced by encoding interference. Further refining my definition of feature types given the evidence presented above, the linguistic information used in dependency resolution will be referred to as a *relevant feature*. This definition then generates a clear distinction between relevant features and the remaining linguistic information that is associated with a word's feature matrix. The features that are not involved in dependency resolution, such as gender in the discussion of Romance RCs, are henceforth *irrelevant features*.

2.9 Hypotheses and predictions

In this review of the literature, I hope to have provided an overview of the historical background and current discussions surrounding cue-based retrieval models. I also provided an overview of the limited literature regarding encoding interference in the realm of sentence processing and several open questions that must be addressed regarding the parser and the retrieval mechanism. These discussions are not comprehensive but are presented to set the stage for the hypotheses that will be tested in the upcoming chapters. Prior to delving into the main hypotheses, I would like to reiterate a key assumption that was presented in the last section: the retrieval mechanism is influenced by external factors that may result in varying strengths/weights of cues. In the case of this thesis, these external factors may be structural, information structural, or even focus-driven in nature, but I mention this underlying assumption because the hypotheses presented in this section make clear predictions about how and where these external factors should generate interference effects.

To further clarify what I mean by external factors influencing retrieval, I would like to briefly discuss two patterns that have been observed with respect to relative clauses and sluices. First off, both of these structures have been shown to possess distinct and independently motivated processing profiles when the position of the target is altered. In the RCs, object extraction has been shown to produce to increase processing delays when compared to subject extraction (e.g. Gordon et al., 2001). Sluices, on the other hand, become more challenging to process when the intended target (correlate) is the subject of the sentence, as opposed to the object; a phenomenon that is known as the Locality bias (e.g., Carlson et al. 2016; Frazier and Clifton, 1998; Harris, 2015; Carlson et al. 2016). For both of these constructions, there is some evidence that the dispreferred target position and respective structure increase the effects of similarity-based interference between the target and the distractor (e.g., Gordon et al., 2011; Harris et al., 2015, 2019).

These findings relate back to my initial assumption that external factors can guide the behavior of the retrieval mechanism, albeit for independent reasons. Although the experimental chapters contain detailed discussion about why these constructions have preferred and dispreferred order in online processing, any evidence of an interaction between the target's position and feature overlap can be used as evidence that the retrieval mechanism takes many factors into account during the selection of an antecedent. On the other hand, if an interaction is not found and only two main effects appear for the presence of the dispreferred structure and for feature overlap between the target/distractor, then we would have some evidence that the retrieval mechanism treats all linguistic information in the same way and is not constrained by external factors when selecting an antecedent. I will return to this discussion regarding structural and unconstrained accounts of retrieval in Chapter 4. Either way, the presence of any target position effect (slowdowns for object-extracted ORCs and for subject-correlate sluices), would, at minimum, be

sufficient evidence that the experimental methods used in this study were sensitive enough to pick up on subtle processing asymmetries.

With the underlying relationship between the two experiments laid out, I will provide a brief overview of the manipulations in Chapters 3 and 4. I conducted two sets of experiments that test the behavior of the morphosyntactic features of gender and number during retrieval. In the first set, I explore the behavior of number as a retrieval cue in subject- and object-extracted relative clauses, as well as test for effects of the potential encoding feature of gender in these constructions. As mentioned in the section on encoding interference, number is the only morphosyntactic feature required for subject-verb dependency resolution in Portuguese, since gender is not marked on the verb that emits a retrieval cue for its subject. The second set of experiments examine how these same features behave in sluicing constructions in Portuguese. Unlike the relative clauses, Portuguese sluices allow both number and gender features to be retrieval cues, as they can be overtly expressed on the wh-remnant that requires an antecedent in the unelided TP. Interestingly, the gender feature can be removed from the wh-remnant without affecting the grammaticality of the sentence, which allows for the manipulation of the relevancy of gender the wh-remnant/antecedent dependency.

Overall, I expect to find that overtly expressed retrieval cues that match with the features of more than one item in memory generate similarity-based interference effects and also to find an effect of encoding interference when items share features that are irrelevant to the retrieval process. Moreover, if the retrieval mechanism indeed considers structural or informational preferences in the selection of an antecedent, the interference effects should be stronger for the dispreferred target location (interaction between feature match and target position). Below, I have presented the

hypotheses that are addressed throughout this thesis.[5] The first set of hypotheses specifically focuses on the nature of the retrieval mechanism itself and the second set considers how an encoding mechanism could be incorporated into current models of retrieval.

2.8.1 Retrieval Hypotheses

This first set of hypotheses address the underlying nature of the cue-based retrieval mechanism. The crucial difference between these hypotheses is how the retrieval mechanism treats relevant and irrelevant features. Current models of cue-based retrieval most closely align with the first, and strictest, hypothesis: relevant feature retrieval.

***A. Relevant Feature Retrieval**:* Only features that are required by a grammatical dependency (relevant features) participate in the cue-matching procedure. If multiple items contain activated relevant features, similarity-based interference effects will surface. Irrelevant features and their respective activation levels are not considered by the retrieval mechanism.

Predictions: In the upcoming experiments, number is always a relevant feature, since it is overtly expressed on the verb in subject-verb dependencies and on the wh-remnant in sluices. When the target and distractor both overlap with the number cue at the retrieval site, interference effects will occur. This logic also applies to cases where gender is a relevant feature (overtly expressed at the retrieval site). This also predicts that gender, which is typically an irrelevant feature (not expressed on the verb, for example), should not produce interference effects. The retrieval cue does not consider the irrelevant feature during the cue-target pairing procedure. Moreover, any interference effects should only appear at the retrieval site.

A slightly more flexible hypothesis is that of *cue super-additivity*. This hypothesis assumes that the retrieval mechanism is primarily sensitive to relevant features, as those provide direct

[5] Experiment-specific hypotheses and predictions will be presented in their respective chapters.

matches between the cue and its target; however, the competition generated by irrelevant feature matches can also result in interference effects in some cases. This hypothesis could potentially explain an interaction between relevant and irrelevant feature match, without having to consider the incorporation of a secondary mechanism to account for irrelevant interference effects (e.g., an encoding mechanism).

B. Cue (Super)-Additivity: All features that have been encoded with an item in memory are capable of influencing the retrieval process to varying extents. The retrieval mechanism first considers relevant features in the cue-matching procedure. If items contain overlapping relevant features that satisfy the retrieval cue, similarity-based interference effects will surface. However, irrelevant feature matches can strengthen the effects of relevant feature interference by increasing the overall activation levels of the target and distractor.

Predictions: As in the relevant feature hypothesis, relevant feature matches will generate similarity-based interference when they overlap. What makes this hypothesis unique is the prediction that the interference effects caused by relevant features will be strengthened if irrelevant features also match. This is described as super-additive for two reasons. First, it is not predicted that irrelevant features will cause interference on their own. This prevents this hypothesis from being "additive," since the effect should not equal the sum of irrelevant + relevant interference. Second, the effect when both irrelevant and relevant features match should be greater than either effect individually, which is what makes this effect super additive.

Finally, it is possible to imagine a null hypothesis for both (A) and (B), which I will refer to as *generalized feature retrieval*. Both of the hypotheses above predict that relevant and irrelevant features should present unique interference profiles; however, it is a logical possibility that they could be identical. It is highly unlikely that this null hypothesis can account for the nature

of the retrieval mechanism in light of the experiments demonstrating cue-weighting in the previous section. In particular, the fact that some cues only showed interference effects when other cues were also matching (discussion in Parker, 2019). However, there will be points where this hypothesis and its predictions are mentioned throughout the dissertation.

C. Generalized Feature Retrieval: The retrieval mechanism is sensitive to all features that overlap between potential antecedents, regardless of their relevancy to the grammatical dependency at hand.

Predictions: The retrieval cue initially matches with its intended antecedent, but interference effects can arise from any other features shared between items in memory. The source of all processing slowdowns is feature overlap between items in memory *at the site of retrieval*; even the features aren't expressed by the retrieval cue (e.g., Portuguese verbs being sensitive to gender even though they aren't morphologically marked for this feature). This would generate the prediction that we would observe no difference between gender and number in any experiment, regardless of how relevant they are for grammatical dependency resolution.

2.8.2 Encoding Hypothesis

Suggesting the need for a second mechanism to account for interference entails a substantial amendment to current cue-based retrieval theory. Although a model with both an encoding and retrieval mechanism would be less parsimonious than making a more "comprehensive" retrieval model that can account for irrelevant feature interference, there are definite empirical and theoretical benefits to making this change. First and foremost, encoding interference, or at least interference that can't be explained by retrieval, has been found in several studies focusing on cue-based retrieval. The current models fail to account for interference from irrelevant features, so adding a mechanism that explains these effects would generate a more realistic account of

similarity-based interference in sentence processing (discussion in Villata et al., 2018). By including this mechanism, we could easily explain how irrelevant features influence activation levels before the retrieval process, whereby showing independent effects of interference for irrelevant features and perhaps stronger retrieval interference effects. Although the encoding mechanism could reasonably interact with any of the proposed retrieval mechanisms above, I will present only the strictest hypothesis for discussion. This is because exploring all of the possible combinations of retrieval and encoding mechanisms would generate highly overlapping predictions. It is more experimentally prudent to test the strictest hypothesis with the most discernable predictions, which can be investigated in more detail in future research.

(C) Encoding Interference: Processing slowdowns may also occur purely as a result of lexical feature matches (those not required for grammatical dependency resolution), as feature overwrite during encoding ultimately degrades the memory representations of both the target and distractor. These effects may appear before or at the retrieval site, but a main effect of feature match should always appear for irrelevant features.

Predictions: Processing delays will still be observed even when the feature is not related to the retrieval cue. In other words, irrelevant features should produce independent effects of similarity-based interference. These effects should surface in the form of a main effect of irrelevant feature matches, and ideally, these effects should surface before the retrieval site. This story becomes more complicated, however, because encoding interference effects also directly influence the features involved in retrieval. This means that encoding interference effects could also appear at the retrieval site for both the relevant and irrelevant features. Moreover, these combined effects at the retrieval site could result in an interaction, which overlaps with Hypothesis B. The crucial difference between this hypothesis and Hypothesis B is that a main effect of irrelevant feature

match is always predicted. As a reminder, Hypothesis B states that irrelevant features should only contribute to similarity-based interference when the relevant feature also matches. Moreover, if interference effects are observed prior to encountering a retrieval cue, there is evidence that a non-retrieval interference is occurring. This would predict in the upcoming experiments that gender, when not overtly expressed as a retrieval cue, will consistently cause reading time delays and/or decrease comprehension question accuracy (as discussed in studies by Villata et al., 2018 and Laurinavichyute et al., 2017).

With the hypotheses outlined, it is important to relate them back to our current understanding of how memory interacts with sentence processing. The three retrieval hypotheses most closely align with the current models of content-addressable memory, where a retrieval cue has direct access to its target via its content. A retrieval cue will be paired with its target in memory by directly access the target based on a subset of its features that overlap with the cue. However, the three hypotheses generate distinct predictions about how much content will be accessed during the retrieval procedure. In particular, a potential issue with adopting either the (Super) Additivity or the Generalized Feature mechanisms is that a greater amount of content is inherently going to be activated in memory and involved in retrieval because the cue to will match with a greater number of competitors. This could mean that multi-cue/feature retrieval models would experience more interference effects than the Relevant Feature Retrieval hypothesis, thus making them less effective.

Another concern with extending the retrieval mechanism to account for effects of irrelevant features, which is not done in most current models, is that this expanded model might not actually represent the psychological reality of the retrieval mechanism and its trans-domain functions. Although incorporating a second, highly specified encoding mechanism introduces greater

theoretical complexity, there is a benefit in proposing two discrete mechanisms that can account for the effects of irrelevant features that can be tested in a variety of cognitive domains. These two mechanisms could also be studied in non-language domains in order to make more domain-general claims about the mechanisms involved in storing and retrieving items from memory. By finding independent motivation for an encoding and a retrieval mechanism and clearly defining their functions within memory, researchers could potentially can avoid the risk of over-expending the retrieval mechanism to account for effects that it does not generate. Although this dissertation cannot make concrete claims in favor of one mechanism over another, I hope to present clear predictions about what a retrieval only or a retrieval + encoding mechanism would look like in sentence processing, in the hope that these topics continued to be researched in the future.

Chapter 3: Relative Clauses

3.1 Review of the Relative Clause Literature

Given the breadth of the research questions, I have elected to use relative clauses as one of the syntactic constructions that will be analyzed in the experiments on Portuguese. Relative clause constructions have been highly studied due in online processing to better address questions such as ambiguity resolution, the parsing of underlying structures, the syntax-prosody interface, and both first- and second-language acquisition. Although retrieval appears to be involved in the processing of most grammatical dependencies, the vast literature on the syntax and processing of relative clauses will provide a structured base from which I can address novel questions about cue hierarchies, the nature of the retrieval mechanism, and potentially explore the concept of encoding interference. Moreover, relative clauses contain a well-documented processing bias, where object-extracted relative clauses obtain greater processing penalties that subject-extracted relative clauses. Importantly, this asymmetry has been shown to modulate similarity-based interference, strongly suggesting an interaction between processing preferences and feature-driven retrieval. As mentioned in the previous chapter, the manner with which the retrieval mechanism is constrained by processing biases is a link between relative clauses and other constructions, like the sluices discussed in Chapter 4. I organize the review of the literature in this section to present the syntactic, processing, and memory arguments that have been used to account for the RC asymmetry that inspired the design of the experiments.

3.2 The syntax of relative clauses

In this dissertation, I conducted three experiments using restrictive relative clauses (henceforth RRCs) in Portuguese (English examples in 3-5, adapted from Sportiche et al., 2013). Although this thesis does not focus on developing the syntax of these structures, a thorough presentation of

how they are formed provides a more linguistic background to discussions of retrieval that are more present in psychology. Andrews (2007) defined RRCs as "a subordinate clause which delimits the reference of an NP by specifying the role of the referent of that NP in the situation described by the RC." In other words, RRCs are used to provide specific information about the noun they modify and distinguish it from other potential referents in the discourse. The underlying structure of RRCs has been heavily debated in the literature, primarily regarding the syntactic position from which the head noun originated (RC internal vs. RC external). Since most of the analyses of relative clause structure have been conducted for English, I will present the English-based arguments and then discuss how these analyses have been applied to Spanish and Portuguese.

(1) Jerry was admiring to [$_{DP}$ the [$_{NP}$ artist [$_{CP}$ *Op* who $_{i}$ [$_{IP}$ I met _____ yesterday]]]].
(2) Beth bought [$_{DP}$ the [$_{NP}$ jacket $_{i}$ [$_{CP}$ *Op* that $_{i}$ [$_{IP}$ I wanted _____]]]].
(3) Katie wanted [$_{DP}$ the [$_{NP}$ cellphone $_{i}$ [$_{CP}$ *Op* $_{i}$ [$_{IP}$ Maria bought _____]]]].

I would like to begin by discussing the internal structure of relative clauses, without referring to the origin of the head NP. As shown in the examples below, English relative clauses contain a wh-phrase, complementizer or phonologically null operator in Spec CP, as seen in the examples above (adapted from Sportiche et al., 2013). Unlike English, Portuguese RCs must contain phonologically expressed wh-expressions such as *qual* – "which", *quem* – "who" (to name a few) or the complementizer *que*/*que* "that".

(4) a. [$_{NP}$ O chocolate$_{i}$ [$_{CP}$ *Op* $_{i}$ que [eu comi t$_{i}$ ontem]]...
The chocolate that I ate yesterday...
b.* [$_{NP}$ O chocolate$_{i}$ [$_{CP}$ *Op* $_{i}$ [eu comi t$_{i}$ ontem]]
*The chocolate I ate yesterday...

I assume that the relative operator, which may be overt or silent in English, is generated in an argument position within the IP and moved to Spec CP position via A' movement (Bhatt, 2002;

[6] The index $_{i}$ means that the elements are coindexed

Chomsky, 1977; Kennedy, 2014). The similarities between wh- and relative clause movement provide convincing evidence that A' movement is required in these constructions. In particular, relative clauses are subject to the same constraints as wh-questions, which were originally outlined in Ross (1969), and the data below are provided to exemplify this claim (adapted from Sportiche et al., 2013 and Santorini et al., 2007, see Sportiche et al. 2013 for a definitions of these constraints). In all of the examples below, the ungrammaticality derives from extracting a wh-word form an illicit site, which can be identified by the indexation of the traces [t] and their coindexed wh-expression marked by the subscript [X_i]. The case that is most relevant for the structure of RCs can be observed in (9), where the constituents must remain in situ if they are derived from within the RC itself.

Wh-Island Constraint
(5) *The [man]$_i$ [$_{CP}$ who$_i$ [$_{TP}$ I wondered [$_{CP}$ who$_j$ [$_{TP}$ I introduced t_i to t_j]]]].

Noun Complement Constraint
(6) *The [woman]$_i$ [$_{CP}$ who$_i$ [$_{TP}$ I belived [$_{DP}$ the claim [$_{CP}$ that [$_{TP}$ I introduced t_i to you]]]]].

Relative Clause Constraint
(7) *The woman$_i$ [$_{C\{}$ who$_i$ [$_{TP}$ I dislike the journalist [$_{RC}$ who interviewed t_i for the talk show]]]

Complex NP Constraint (relative clauses) – A constituent cannot be removed from within a relative clause (a complex NP)
(8) *The [man]$_I$ [$_{CP}$ [$_{TP}$ I like the [woman]$_j$ [$_{CP}$ who$_j$ [$_{TP}$ I introduced t_i to t_j]]]].

Sentential Subject Constraint
(9) *The [addiction]$_i$ [$_{CP}$ which [$_{CP}$ {that he admitted/ admitting} t_i] nearly destroyed his career]]]

Left Branch Constraint
(10) Mary loves * the [woman]$_i$[$_{CP}$ whose$_i$ [$_{TP}$ you met t_i friend]]].

Coordinate Structure Constraint
(11) Sara disliked *[the dessert]$_i$ [$_{CP}$ which$_i$ [$_{TP}$ you ordered {coffee and t_i, t_i and coffee}]]]

Although operator movement seems relatively straight-forward, the syntactic relationship between the RC and the NP that it modifies has been heavily debated in the syntactic literature. There are two primary claims about the external NP head of a relative clause: 1) the NP originates in an RC external position or 2) the NP is moved from an RC internal position. Bhatt (2002) further divides these categories into three separate hypotheses, the external head hypotheses, the raising hypothesis and the matching hypothesis, which are outlined below.

One of the earliest analyses for RC structure is the head external analysis (Chomsky, 1977; Partee 1975, etc.). This analysis claims that the head NP originates from a position outside of the relative clause and undergoes movement to arrive in its surface position, without moving into relative clause (see Tree 1). Within the relative clause, there is A'-movement of a relative operator (either overt or phonologically null), which occupies Spec CP position in the final structure. This interpretation of the sentence is driven by co-indexation of the RC operator and the trace of the RC external NP. (Example and discussion in Bhatt, 2002).

(1) the book [$_{CP}$ Op$_i$/which$_i$ John likes t$_i$]

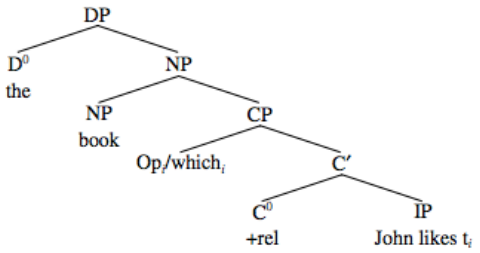


***Figure 2**: Externally Headed Relative Clauses (Bhatt, 2002)*

The second analysis for RC structure is the head raising analysis (Brame, 1968; Schachter, 1973; Kenedy, 2014 among others). This analysis assumes that head NP originates an RC internal position and to its final surface position. This is the first of two hypotheses that reference internally headed relative clause structures. Crucially, the head NP is only interpreted in its relative clause internal trace position (Hulsey and Sauerland, 2006). Since the head originated within the relative

clause position, reconstruction is possible, and the head can be interpreted in an RC internal position (analysis and example from Bhatt, 2002).

(2) the [book]$_j$ [$_{CP}$ [Op/which t_j]$_i$ John likes t_i]

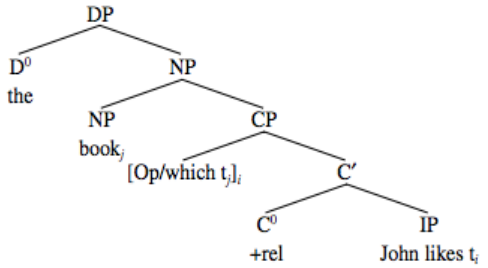


Figure 3: *Raising Hypothesis for Relative Clauses (Bhatt, 2002)*

There is syntactic evidence from binding and subcategorization that a raising analysis is supported in English relative clauses, but some of the simplest evidence comes from idioms (see Bhatt, 2002). Bhatt (2002) adapts previous data (e.g. Marantz, 1984) to show that an idiom can only exist as part of a larger phrase *or* as the head NP of a relative clause when the trace of the NP pertains to the larger expression that constitutes the idiom. If an external head analysis were true, there is no reason that the (b) cases in the examples below should be ungrammatical, as constituents should be able to be removed from idioms and placed elsewhere. The fact constituents cannot be removed from idiomatic expressions implies that a trace of the moved element must remain in the original position of the constituent, and this is trace is relative clause internal in the structures below.

(12) (attributed to Brame 1968; ex. (35) from Schachter 1973 in Bhatt, 2002)
a. I made headway.
b. *(The) headway was satisfactory.
c. The headway that $_t$ I made was satisfactory (Schachter, 1973).

(13) (attributed to George Bedell; fn. 15 from Schachter 1973 in Bhatt, 2002)
a. He solved the problem in a clever way.
b. *The clever way impressed me.
c. The clever way in which he solved the problem $_t$ impressed me.

An alternative hypothesis for internally headed RCs is the Matching Hypothesis (Bhatt, 2002; Harris, 2008; Lees, 1960; Chomsky, 1965; Sauerland, 2004). The matching approach

essentially combines the approaches in the external head and head raising analyses. This hypothesis operates under the assumption that there is an external head noun as well as an RC internal noun that is phonologically deleted before production. The head NP is interpreted outside of the relative clause and a copy (or an NP similar enough to license ellipsis) is interpreted RC internally (Hulsey and Sauerland, 2006; Sauerland, 2004). Distinct from the other analyses, there is no syntactic movement, as shown in the example below from Bhatt (2002).

(3) the [book] [$_{CP}$ [Op/which book]$_i$ John likes t$_i$]

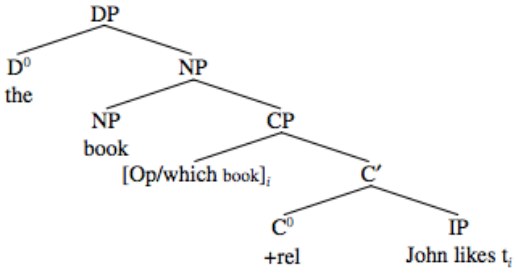


Figure 4: *Matching Hypothesis for Relative Clauses (Bhatt, 2002)*

Choosing between these three analyses is not straightforward, as there is evidence cross-linguistic that grammars permit more than one type of relative clause structure (Bhatt, 2002 for an overview; for languages with both types of RCs - Aldrige, 2017 for Tagalog, Cole 1978 for Quecha; Chung and Kim, 2002 for Korean among others). With that being said, recent work on English, Portuguese and Spanish relative clause structures have found more evidence for internally headed RCs than externally headed RCs (Bhatt, 2002, Sauerland, 2003 for English; Kato and Nunes, 2009, Kenedy, 2002 for Portuguese; Arregi, 1998, Plann, 1975, 1980 for Spanish). For this reason, I assume an internally headed RC structure for the sentences that will be analyzed in this experiment, as this has received the most evidence in the literature and allows for cross-linguistic comparisons between the languages looked at in this dissertation. However, I will not choose

between a Raising or Matching approach, as they generate similar predictions for the experiments discussed in the following sections. For my purposes, the exact underlying structure of the RC should have a minimal effect on how retrieval cues affect the processing of these sentences. Aside from a cue-based retrieval mechanism, I also I assume an incremental parsing mechanism that, upon encountering a relative pronoun or complementizer in its surface position, makes the decision to parse the sentence as a relative clause and attempts to fill the most local gap site (described in Frazier et al., 1983); therefore, the exact syntactic derivations involved in forming this structure are not particularly relevant to the questions that I am asking.

3.3 Relative Clauses in Sentence Processing

Aside from the theoretical debates about their underlying structure, relative clauses are also one of the most well-studied constructions in sentence processing. In particular, the processing literature has focused extensively on identifying why object-extracted relative clauses (ORCs) are more difficult to process than subject-extract relative clauses (SRCs). Examples of an SRC and an ORC are given in (14a) and (14b), respectively.

(14) a. [The man$_i$ who ____ hated the sheriff ____$_i$] $_i$ went to jail.

b. [The man$_i$ who the sheriff hated _____$_i$] $_i$ _____ $_i$ went to jail.

The experimental evidence for the processing asymmetry between SRCs and ORCs is copious and has been observed in several languages using a variety of methods. Although this list is not exhaustive, from the methodological standpoint, the object-extracted RC penalty has been observed in self-paced reading (Gordon et al., 2001, 2004; Myamoto and Naknamura, 2003), eye-tracking (Betancort et al., 2009; Traxer et al., 2002), and ERP (King and Kutas, 1995; Ueno and Garnsey, 2008; Weckerely and Kutas, 1999). Cross-linguistically, this asymmetry has been identified in Brazilian Portuguese (Gouvea, 2003), English (Traxler et al., 2002), Dutch (Frazier,

1987; Mak et al., 2006), French (Holmes and O'Regan, 1981), German (Schriefers et al., 1995), Greek (Stavrakaki, 2001), Hungarian (MacWhinney and Pleh, 1988), Japanese (Ishizuka et al., 2006; Miyamoto et al., 2003; Ueno and Garnsey, 2008), and Spanish (Betancort et al., 2009) to name a few. Finally, this phenomenon is observed for both L1 acquisition (Belletti et al., 2012; Sheldon, 1974) and L2 acquisition (Eckman et al., 1988; Doughty et al., 1991).

3.4 Processing explanations for the asymmetry

The theoretical accounts for this processing asymmetry can be classified into three main categories: memory accounts, parsing accounts, and surprisal accounts. In the following sections, I will provide a brief overview of the primary claims made in each of these categories. Although this dissertation focuses on retrieval as one of the main factors in this asymmetry and object-extraction as a whole (following Gordon et al., 2001, 2002; Doughty et al., 1991) I present these alternatives to show that there may be more than one reason that ORCs are harder to process. I focus primarily on the retrieval account of the asymmetry due to large amount of empirical support, but I designed my experiments using ORCs because they are sensitive to interference effects more than similar structures (SRCs, for example). This sensitivity makes ORCs ideal testing grounds for both well-known and understudied cues, such as number and gender.

3.4.1 Memory accounts

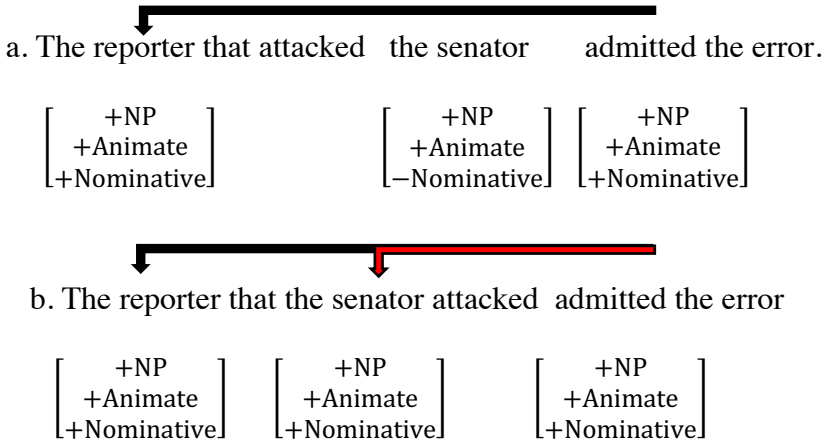
Some of the most prevalent accounts of the SRC/ORC asymmetry in the recent literature are related to memory mechanisms. Within the memory accounts used to explain the ORC penalty, I revisit the debate between the role of decay and interference that was presented in Chapter 2, Section 2.2.3. I will refer to the first category of hypotheses as storage-based accounts (Miller and Chomsky, 1963; Ford, 1983; Frazier & Fodor, 1978; Gibson, 1998; Lewis, 1966; MacWhinney, 1987; Wanner & Maratsos, 1978). The accounts are most commonly represented by the

Dependency Locality Theory, which claims that processing costs are driven by integration (attachment) and storage in memory (Gibson, 1998, 2000). This theory predicts that object-RCs are harder to process than subject-RCs because the filler phrase must be stored in memory as an unattached constituent for a longer period of time in ORCs than in the SRCs. As time passes, the trace of the filler phrase decays from memory and the sentence becomes harder to process. A modified version of this hypothesis was experimentally tested by Wanner and Maratsos (1978), who found that comprehension question accuracy and name recall were diminished when subjects were presented with a list of words between the filler item and the gap than if they were presented before the filler. The authors attribute these results to limited working memory resources, which are used up when the comprehender is storing the filler in memory and simultaneously processing the intervening information. Since there is not enough space within WM to store the filler when WM resources have reached capacity, retrieval accuracy decreases as the trace of the filler undergoes temporal decay due to lack of activation. However, as discussed in sections 2.3.2 and 2.4, the empirical evidence that supports a purely decay-based approach to sentence processing has been shown to be primarily task-driven and incomplete.

Alternatively, similarity-based interference accounts have been used to explain the ORC/SRC asymmetry. Unlike the storage accounts, these approaches state that ORCs are harder to process than SRCs because upon arriving at the matrix verb, which contains the retrieval cue, the comprehender must retrieve the matrix subject over the intervening RC subject, which both possess nominative Case and other overlapping features. The influence of Case in interference explanations strongly suggests that retrieval must proceed an initial syntactic parse that generates these features. In the example below, this means that the comprehender must retrieve *the reporter* in (15b) over the intervening information *the senator*. Given that both the matrix subject and the

RC subject can potentially satisfy the retrieval cue's associative process, there is increased interference for ORCs.

(15)



Further support that this asymmetry is driven by retrieval comes from studies in which degree of semantic overlap between the target and distractor were manipulated (Gordon et al. 2001, 2002, 2005; Van Dyke and McElree, 2006). Gordon et al. (2002) asked participants to memorize a 3-word list and then read a sentence. The experiment manipulated the similarity between the nouns in the memorized list and the sentential nouns to observe whether SBI modulates the ORC processing cost. The results showed that subjects responded less accurately to comprehension questions when the items in the memorized list matched the type of the sentential target noun. For example, referring to (15), participants may have been asked to memorize three common nouns to match *the senator* (e.g. baker, lawyer, doctor) or three names that do not match the sentential NPs (e.g. Bob, Mary, and Sue). Additionally, Gordon and colleagues corroborated previous work showing inflated reading times on the main verb in ORCs in cases where the was a semantic match between the memory-load list and the sentence. Most relevant for the SBI account of this asymmetry is that the processing penalty for ORCs was nearly eliminated in cases where there was a semantic mismatch between the target and the distractor, strongly suggesting that the asymmetry

is merely a byproduct of similarity-based interference. In the previous studies reported for the processing asymmetry, the semantic similarity between the target and distractor was not manipulated and typically pertained to the same semantic category (i.e. banker and lawyer), which presents a confound in the results and conclusions drawn from them.

To further investigate whether Gordon et al.'s (2002) findings supported encoding interference or proactive interference, Van Dyke and McElree (2006) a follow-up study. The authors controlled the encoding information by using a consistent memory-load list and only manipulated the semantic relationship between the memory load items, the intended subject of the sentence and the verb. For example, subjects were asked to memorize "TABLE – SINK -TRUCK" and were presented with a sentence like "It was the *boat* that the guy who lived by the sea *sailed/fixed* after 2 days." When the sentence contained the verb sailed, only the intended subject was semantically related to the verb; however, when the verb *fixed* was used, all of the memory load items, as well as the intended subject, were semantically related to the verb (see Van Dyke and McElree 2006 for further examples). The results showed slow reading times when participants were presented with a memory load, as opposed to having no memory load. Additionally, retrieval difficulty increased when the memory load items *matched* cues for the verb, but not when the memory load items *did not match* the semantic cues projected by the verb. The authors attribute the results to proactive interference, as opposed to encoding interference because the encoded information remained constant throughout the experiment. However, they did not investigate semantic relatedness of memory load items and cannot completely rule out any effect of encoding interference in their results. Overall, there is a large body of evidence showing that similarity-based interference effects play a role in successful sentence processing and can modulate the difficulty associated with SRC and ORC processing.

3.4.2 Parsing accounts

In addition to the memory accounts, syntactic parsing preferences and the mechanisms that guide them have been used to explain the asymmetry. The first set of parsing analyses can be categorized as "syntactic" accounts (see Traxler et al., 2002 for an overview). These accounts claim that ORCs are harder to process because the filler must change syntactic and semantic roles in ORCs but not in SRCs. In SRCs the filler serves as the subject of both the RC and of the matrix clause, therefore maintaining only one syntactic function and one thematic role throughout the parsing of the sentence. On the other hand, in ORCs the filler is the direct object of the relative clause and the subject of the matrix sentence, therefore the parser must assign two thematic roles and syntactic features to the same element.

Some of the prominent accounts that align with thematic role assignment and/or synaptic approaches are the parallel function account (Sheldon, 1974), the double function hypothesis (Bever, 1974), and the perspective maintenance hypothesis (MacWhinney, 1997). All three of these hypotheses assume that processing difficulty is associated with syntactic and semantic role shifting, but they vary slightly. The Parallel Function account claims that if two coreferential NPs have the same grammatical role, the sentence will be easier to process than one containing coreferential NPs with different grammatical functions (Sheldon, 1974). The Double Function Hypothesis claims a stimulus cannot be perceived as having two grammatical functions simultaneously; therefore, an NP in an RC cannot be interpreting as both the sentential subject and RC object on the first parse, which makes ORCs more challenging to process (Bever, 1974). The Perspective Maintenance Hypothesis states that comprehenders prefer to maintain the interpretation of a sentence that most closely aligns with the individual's perspective/relationship with the interpretation. In other words, humans prefer to adopt an agent perspective for a sentence,

as opposed to a passive recipient perspective because that is how they perceive their role as an interpreter; therefore, when they must adopt the dispreferred perspective in the case of an ORC, they find this sentence more difficult to process (MacWhinney, 1977). The Perspective Maintenance Hypothesis is not particularly grounded in current models parsing and has received very little follow-up in the literature, so it will be put aside.

These types of accounts have been largely discredited by several experimental studies. A recent study by Traxler et al., (2002) showed that adults had no difficulty simultaneously interpreting an inanimate sentential subject as both the subject of the matrix clause and object of the relative clause, implying neither the Parallel Function nor the Double Function hypotheses completely account for the ORC/SRC asymmetry in adult sentence processing. Essentially, the subjects did not experience processing difficulty when they were required to shift the thematic role of an NP in the sentence, which is what the syntactic approaches would predict. Additionally, cross-linguistic research on Hebrew (Schlesinger, 1975) and Japanese (Hakuta, 1981) have shown that the double function accounts are not supported in languages with flexible word order because an account dependent on grammatical function alone should no effects of word order; however, in these studies, there is an interaction between construction type (word order) and the ease of processing the ORC (discussed in Lewis and Vasishth, 2005).

An alternative parsing hypothesis is the *Active Filler Strategy,* which claims that the preference for SRCs is due to the immediate attempt by the parser fill the most local gap position (Clifton & Frazier, 1989; Frazier & Clifton, 1989; Frazier & Flores d'Arcais, 1989; Fodor, 1978; Stowe, 1986 among others). Early versions of this account operated under the assumptions that 1) the parser is a serial mechanism that prefers the least cognitively-taxing parse for initial sentence interpretation, which must be revised when the initial parse does not align with the intended

interpretation of the sentence and 2) limited memory resources influence the decision o f the parser when encountering a filler during the initial parse (Frazier, 1987). As previously discussed, there have been several memory-related hypotheses that suggest that the longer that the comprehender must hold onto an unresolved dependency, the harder that the sentence is to process (due to limited memory capacity). Wanner and Maratsos (1978), which is referenced in Frazier's early formation of the Active Filler Hypothesis (defined below), refer to the role of limited memory resources on sentence processing as the HOLD hypothesis. The initial Active Filler Hypothesis was then defined as: "Empty HOLD as soon as possible" (Frazier, 1987). In other words, upon encountering a filler, the parser will relieve the memory load associated with storing the filler by assigning it to the most local gap site. However, this alone does not explain how to parser mistakenly interprets the sentential subject as the relative clause subject in object RCs because underlyingly, object RCs do not contain a true gap in the subject position.

The active filler hypothesis was further refined to state: "When a filler has been identified, rank the option of assigning it to a gap above all other options" (Frazier and Clifton, 1989). Under this definition, upon identifying a filler, the parser will *predict* a gap in every grammatically legal position and assign the filler to this position. Therefore, in the case of relative clauses, when the parser encounters a filler and the relative pronoun, it postulates a gap in the subject position of the RC and automatically inserts the filler. This rapid parsing strategy is what contributes to the preference for subject RCs and the perceived facilitation effects associated with this structure in online processing. The penalty associated with ORCs is due to a reanalysis cost, as the parser must reanalyze the placement of the filler when it encounters the actual syntactic gap in the object position of the RC. Evidence for the Active Filler Hypothesis has been used as an explanation for several filler-gap dependencies processing preferences, an overview of which is presented in

Frazier and Clifton, 1989. Crucially, hypotheses of this type highlight the interaction between memory systems and sentence parsing, which can be applied to models such as encoding and retrieval.

3.4.3 Surprisal

Surprisal accounts of sentence processing are gaining popularity in the fields of psycholinguistics and computational linguistics (Hale 2001; Levy, 2005, 2008). Unlike the parsing and memory accounts discussed above, the surprisal account states that a comprehender uses probabilistic knowledge of their language, acquired through experience, to generate expectations about the interpretation of the structure that they have already received and the information that is yet to come (Levy and Gibson, 2013). Moreover, these expectations can be quantified and used to predict comprehension difficulty and probable interpretation (Levy and Gibson, 2013). From a more abstract perspective, surprisal is an information theoretical metric that is used to gauge the complexity of a given construction within a grammar. The surprisal metric reflects the probability of incoming linguistic information, where a low surprisal value corresponds to highly probable information and a high surprisal value corresponds to improbable information. Crucially, each incoming word is treated as its own unique event with its own surprisal value, and this serves as a linking hypothesis between the complexity of the sentence and real-time measures of sentence comprehension, such as reading times (Hale, 200l; Levy, 2005, 2008; see Hunter in progress for a description of the mathematical properties of surprisal).

The surprisal hypothesis has been applied to the SRC/ORC asymmetry. Given that SRCs are more common across corpora, at least for English, I can assume that these are more probable structures and should incur a lower surprisal metric (Levy et al., 2013). Therefore, when applying a linking hypothesis between reading behavior and surprisal, the reading times associated with

SRCs should be faster, as these are the expected structure and have a higher probability of occurrence. Importantly, surprisal accounts predict that the comprehender will show reading delays upon encountering the RC subject, which will immediately discredit the expected SRC structure. The problem with this analysis, as applied to the ORC/SRC asymmetry, is that the bulk of the literature on this topic has shown the reading delay to occur on the RC or matrix *verb*, not on the subject of the RC (see Grodner and Gibson, 2005; Levy, 2008; Levy and Gibson, 2013). Staub (2010) addressed these questions by conducting a word-by-word eye-tracking analysis of SRCs and ORCs. As expected, all three reported eye-tracking measures showed reading disruptions on the RC verb (not the matrix verb); however, he did find a slight reading time penalty on the subject in an eye-tracking study. The reading time disruption on the RC subject is predicted by surprisal accounts, but he claims the majority of the penalty was still observed on the verb, which implies that surprisal does not make an entirely accurate prediction for processing relative clauses.

3.4.4 Where does this leave us?

The parsing and memory accounts appear to make more accurate predictions than surprisal when explaining the SRC/ORC asymmetry. Although this dissertation does not focus on teasing apart the theoretical explanations regarding why ORCs are more difficult to process than SRCs, I agree with the stance taken by Levy and Gibson in their 2013 commentary on this topic:

> "I consider theories of syntactic processing making reference to explicit, costly (and/or potentially fallible) memory operations, such as those of Gibson (1998, 2000) and Lewis and colleagues (2005, 2006), of continued importance in the study of RC comprehension because they make the right predictions not only about *what* is difficult but about *where* the difficulty is observed in this heavily studied empirical domain."

For this reason, I will not consider surprisal to be the correct theoretical assumption behind this processing asymmetry, as it makes incorrect predictions about the location of processing

difficulty in ORCs. This is not to say that surprisal should be completely discounted as a metric of processing difficulty, but I believe that the theory, as currently formulated, cannot sufficiently explain the effects observed in the memory-based accounts of sentence processing. Although it is logically and mathematically possible to incorporate finer-grained linguistic information within surprisal theory, I was unable to find any attempts to do so in the literature. As of now, surprisal theory is limited to lexical prediction alone. Therefore, as it stands, surprisal theory cannot account for similarity-based interference in sentence processing.

Since the role of linguistic features is unclear in current models, no assumptions have been made about how SBI and other memory constraints contribute to surprisal accounts. As I see it, there are two distinct possibilities: 1) encountering a word with a specific feature will prime/increase the possibility of encountering another word with the same feature or 2) no priming effects will be observed because the surprisal metric is not sensitive to low-level linguistic information. If the first case were true, surprisal and SBI accounts would be making opposite predictions about the role of feature overlap in sentence processing.

Within the memory hypotheses, I adopt the claim that similarity-based interference likely plays a significant role in the SRC/ORC distinction, but also acknowledge that other parsing accounts can reasonably explain the phenomenon. As discussed in section 2.3 (and related sub-sections), the evidence supporting storage and decay models of sentence parsing has been challenged by the cue-based models, which produce the same effects once interfering information has been removed. Moreover, the experimental evidence provided by Gordon et al., (2001, 2002) that found that the ORC penalty is severely reduced when there is no semantic interference between the sentential NPs and items stored in memory strongly suggest that this asymmetry is driven in large part by retrieval interference. Further evidence is found in computational studies,

as the reading time effects were replicated in the ACT-R model, which operates using the cue-based model (Lewis and Vasishth, 2005).

With that being said, there is also explanatory power in the Active Filler Strategy. As mentioned in section 3.4.1, the cue-based retrieval model is composed of two stages: initial parsing and retrieval. Although the parsing phase of processing is oftentimes ignored in the retrieval literature, I follow Staub's (2010) proposal that the preference for SRCs can easily be explained by a serial predictive parser, where processing penalties are incurred only when the parser encounters a less probable structure and is forced to reanalyze the sentence. Hypotheses such as the Active Filler Strategy can explain the higher probability of SRCs, as a subject-RC is the least cognitively-taxing parse available during initial stages of processing. Lewis and Vasishth's (2005) model of the retrieval mechanism also supports these claims, as the parsing mechanism in their model reflects the general principles of incremental parsing, so long as the model has space to retain partially built structures in memory. For these reasons and those presented in the previous chapter, I adopt the cue-based retrieval model, while maintaining the assertion that an incremental first-pass parser is at play. Initial parsing mechanisms that generate something like the active filler hypothesis are supported by evidence from a variety of constructions, so it is likely that this parsing process either precedes or operates in lockstep with the retrieval mechanism during sentence comprehension. There is no reason to believe that the parser and retrieval mechanism are inherently one and the same, as assigning a basic structure and retrieving antecedents for filler-gap dependencies are theoretically two very distinct processes, so I will assume that these mechanisms are distinct.

3.5 Motivation for Current Studies

Operating under the assumption that the SRC/ORC asymmetry is a psychological reality, with perhaps more than one source, three self-paced reading experiments were conducted to test the role of morphological features in relative clause processing in Brazilian Portuguese. Given the findings presented by Gordon (2001, 2002), Fedorenko et al. (2006), and Van Dyke and McElree (2006), I decided to test the morphological features of gender and number to see if these features behave similarly to semantic-class cues that have been used as support for cue-based retrieval models. If the RC asymmetry can be fully accounted for using a generalized model of cue-based retrieval, I would expect any legitimate retrieval cue to be able to modulate the RC asymmetry in the same way that semantic class information did in the aforementioned studies.

I also return to the questions of cue-relevancy and additivity that have been presented at several points during this dissertation: does the retrieval of irrelevant cues generate similarity-based interference effects? Can irrelevant cues produce super-additive similarity-based interference effects, when the relevant cues are matching? Are these simply effects of encoding interference? In the RC constructions tested, I was able to investigate these questions because number is the only relevant cue employed in the retrieval of a matrix subject, whereas gender is grammatically required by Portuguese.

3.5.1 Hypotheses and Predictions

This section presents a brief overview of the theoretical hypotheses presented in Chapter 2 and their overarching predictions. The specific predictions that the hypotheses generate for each experiment will be presented alongside the experiments themselves. Unlike the general overview in Chapter 2, this section also presents the primary ways to distinguish between hypotheses that generate similar predictions.

To begin, there are three primary hypotheses for retrieval that I will be addressing. As discussed in Chapter 2, the information that constitutes a retrieval cue and can generate similarity-based interference effects is still an open question in the theory of cue-based retrieval. Following the discussion, I will present a fourth hypothesis that focuses on a second mechanism that could generate processing delays, the encoding mechanism.

The ***Relevant Feature Retrieval*** hypothesis claims that only features that are required for the resolution of a grammatical dependency undergo cue-matching procedures. The concept of a feature being "required" by a dependency will vary on the structure being analyzed. As a reminder, these required features will be referred to as *relevant features*. In the subject-verb dependencies analyzed in this chapter, I assume that the relevant features are those projected by the verb to select a subject.

At minimum, the relevant features for subjecthood in Brazilian Portuguese are [+NP], [+Nominative] and [+/- Plural], where subject must agree in number with the verb itself. All of these features are capable of modulating the intensity of similarity-based interference in relative clauses. Under this approach, the primary distinction between SRCs and ORCs is the fact that ORCs contain two [+Nominative] [+NP] elements, which may or may not match on the feature [+/- Plural]. This predicts that the RC asymmetry is initially generated by the overlap of [+Nominative, +NP] and that interference-driven processing delays would worsen when additional features overlapped (e.g. Number). Therefore, we should observe a consistent main effect for a penalty for ORCs, as well as an interaction between number and clause type. This interaction is important because it demonstrates how a structural feature (Case) is able to modulate the interference caused by other features, demonstrating that the retrieval mechanism is indeed sensitive to both types of features. However, if this interaction does not appear, there are other

potential explanations. The absence of an interaction but two main effects for ORC and Number Match would indicate 1) that the method used was sensitive enough to pick up on the ORC penalty and 2) that number features generate interference in Portuguese, even though these effects could not answer questions about the involvement of structural information in antecedent selection. Importantly, all NPs in Portuguese also contain *irrelevant* gender markers, which I will denote as.[+/- Feminine]. If this and only this hypothesis is supported, there should be no observable effects of gender interference in these experiments. Moreover, this approach aligns with arguments in favor of a cue weighting system where syntactic features are weighted higher than other features (e.g. Van Dyke and McElree, 2007; Parker, 2019). This weight potentially contributes to the ORC itself and explains why number match effects should only surface in the form of an interaction.

In contrast, the hypothesis ***Cue (Super)Additivity*** generates more flexible predictions that the previous hypothesis. Under this approach, the retrieval mechanism is *primarily* sensitive to relevant features but also shows sensitivity to irrelevant features when the relevant features are matching. When both relevant and irrelevant features match, there is a super-additive effect at the retrieval site that would not be observed in a mechanism that treats all cues equally. As mentioned in Chapter 2, there has been a fair amount of discussion about cue combinatorics, as often retrieval cues overlap on more than one feature with the target/distractor. It is likely that super-additive effects will surface. Van Dyke and McElree (2011) claim that there is no additive effect for syntactic and semantic cues (nor super-additive, in their findings), and Dillon et al. (2013) suggest that, in anaphors, non-structural cues are weighted at 0 and point out that this might not be applicable to all structures. In this study, both gender and number are structural, but only number is involved in the dependency. It is unclear how these previous cue-weighting approaches would handle this information. However, if we are able to identify a basic weighting differential between

relevant and irrelevant structural features, it will become much easier for future research to develop a coherent theory of the factors involved in models of cue combinatorics.

The null hypothesis, the ***Generalized Feature Retrieval*** hypothesis states that the retrieval mechanism is sensitive to all features that overlap between potential antecedents, regardless of their relevancy to the grammatical dependency at hand. This hypothesis predicts that both relevant and irrelevant features would produce identical similarity-based interference effects. Therefore, both number and gender features will cause similar levels of interference in the form of an interaction with clause type. Importantly, these effects should arise at the retrieval site itself, which is the matrix verb in the constructions discussed in this chapter. I point this out, as I will revisit the location of irrelevant feature interference effects in the encoding hypothesis.

Stepping away from a purely retrieval-driven approach to interference, the ***Encoding Interference*** hypothesis generates predictions about similarity-based interference that stem from the encoding mechanism itself. This hypothesis postulates that processing slowdowns may also occur purely as a result of lexical feature matches (those not required for grammatical dependency resolution), as encoding similar items results in degraded memory representations. Looking at the retrieval hypotheses above, it is relatively challenging to separate encoding interference from something like the Generalized Feature Retrieval Hypothesis or even the Cue Additivity Hypothesis, as all three approaches predict an effect of irrelevant features. However, there are a few fine-grained details that can separate the three.

To begin, none of the retrieval hypotheses would predict an independent main effect of gender match that is not modulated by clause type information. To spell this out more clearly, the retrieval approaches are all hinged on the fact that ORCs are more challenging than SRCs due to an underlying [+Nominative] match between the matrix and embedded subjects. It is only in cases

where this overlap is present that any additional feature manipulations should give rise to interference effects at the retrieval site. Therefore, encoding interference should primarily surface in the form of a main effect of gender match, independent of clause type, at some point before the retrieval site. As Van Dyke and McElree (2006) pointed out, although encoding effects would minimally appear before the retrieval site, this is not always the case. As discussed in Chapter 2 (Section 2.6), the feature overwriting process responsible for encoding interference may degrade the activation level of the intended target so much that encoding interference effects can surface at the retrieval site itself.

How, then, can one differentiate between the retrieval hypotheses and encoding interference? Ideally, the approaches can be separated by both the location of the effects (before the retrieval site) and the presence of a main effect for irrelevant feature match that does not interact with clause type. I will return to this discussion in more detail in Experiments 2 and 3, as well as in Chapter 4, but for now, any main effect of gender match that appears before the retrieval site and is independent of clause type will be interpreted as encoding interference, since retrieval interference in any form would predict an interaction between clause type and relevant feature match and no simple main effect of relevant feature matches.

3.5.2 General Method

All three self-paced reading experiments were conducted using similar methods and procedures, so I will present a general overview here and provide further details about minor changes between the experiments in their respective sections.

3.5.3 Participants

A total of 32 native Brazilian Portuguese speaking participants were analyzed in each experiment. In each of the studies, participants were removed from the analysis if they did not pass the

screening measures that were used to verify attention to the study and active participation[7]. The controls that were considered for subject exclusion were: (1) above 80% accuracy on filler comprehension questions, (2) above 80% comprehension question accuracy for 5 catch questions that either contained grammatical mistakes or asked questions about common idiomatic expressions of Brazilian Portuguese and (3) participants were asked to self-report that they did not have a reading or attention disorder.

Participants were all associated with the State University of Campinas in Campinas, Brazil and were recruited primarily via posts on social media pages (Facebook) associated with the university and snowball recruitment methods. Participants were offered a ticket for a raffle containing small prizes as compensation and were offered one additional ticket for each participant that they referred to the study.

3.5.4 Materials

Participants saw 20 relative clause sentences in all three experiments. The experimental items were nearly identical across experiments, with the exception of the factors that were manipulated in each experiment. This was done to facilitate cross-experimental comparison. Being able to show that certain trends were replicated in each experiment helps to support our claims in this section, as it shows that distinct groups of subjects show the same reading patterns regardless of the fillers and subexperiments that they were exposed to. The 20 RC items were interspersed between 5 catch items, 24 experimental fillers (including the sluicing items reported in the next chapter), and roughly 80 non-experimental fillers[8]. The high quantity of filler items was intended to distract

[7] Four participants were removed from the Number Study (1), two were removed from the Gender Study (2) and one was removed from the Gender + Number study (3).

[8] Number of non-experimental fillers: Experiment 1 (65) and (16) non-experimental sluices, Experiment 2 (65) and an additional (16) non-experimental sluices, Experiment 3 (65), an additional (12) items used in a sub-experiment on pronoun resolution, and (4) non-experimental sluices. All studies had 130 sentences.

participants from perceiving the sub-experiments, one of which contained a relatively salient structure (sluices) and to maintain a roughly 1:4 ratio between experimental items and fillers. Finally, to verify that all sentences sounded natural and did not contain any grammatical errors, all items and fillers were read by two native speaking research assistants prior to administrating the experiments.

All of the nouns in the experiments were animate nouns representing a profession, which represents the most gender-flexible class of nouns in Portuguese. It is important to note that Portuguese contains a class of "pseudo-neuter" nouns for professions, typically ending with *–ista* that are only disambiguated by the presence of a determiner that is overtly marked for gender (*o jornalista* – "male reporter" vs. *a jornalista* – "female reporter"). None of the nouns used in the experiment pertained to this pseudo-neuter class, as I wanted to guarantee that the gender of the item was morphologically marked on the noun itself. In addition, previous studies on Spanish have found that nouns with overt gender marking (*médico* ("doctor$_{\text{MASC.}}$) vs. *médica* (doctor$_{\text{FEM}}$) are less likely to generate gender bias effects in online studies, further supporting the use of overtly gender-marked nouns (Carreiras et al, 1996). Gender bias in nouns refers to the fact that certain professions are expected to be either masculine or feminine (e.g. nurse is typically feminine) and violating these expectations can lead to processing delays that could be potential confounds in my study. Given that Spanish and Portuguese contain extremely similar morphology and lexicons, I assume that the claims about gender biases in Spanish can also be extended to Portuguese.

Although the gender-marked nouns do not seem to generate bias effects in online studies, the gender bias (which is masculine dominant) is directly observable in the frequency with which a certain gender of a noun appears in online corpora of Portuguese. For this study, I used the *Léxico do Português* corpus (Estivalet, 2015). In order to account for the effects of this bias, I averaged

the frequency and length of both the masculine and feminine versions of the noun in question and used the combined averages for lexical characteristic matching. All targets and distractors were then matched on length and frequency. The targets had an average frequency of 862.45 (SE=384.32) and an average length of 8.57 (SE=0.67), and the distractors had an average frequency of 1035.88 (SE=264.53) and an average length of 8.8 (SE=0.4). A paired *t*-test was used to verify that the lists were not statistically different from one another (length, $t(19) = -0.49$, $p = 0.63$; frequency, $t(19) = -0.33$, $p = 0.74$).

To take further precautions against the gender bias, half of the items in this experiment contained a target in its preferred gender and the other half of the items contained a target with the dispreferred gender. The target remained in the same gender form for all experimental items and only the distractor was manipulated between conditions. This manipulation provides a way to statistically compare participants' behavior for bias-conforming and bias-non-conforming nouns. Unrelatedly, half of the items were plural and half of them were singular, to control for any effects of plurality on the retrieval process. This was done to prevent the subjects from adopting strategies to only look for either singular or plural nouns.

3.5.5 Design

In the first two experiments, a 2 X 2 design was used, crossing Feature Match (Match, Mismatch) and RC Type (SRC, ORC). In the first SPR experiment, the number feature was manipulated between the target (matrix subject) and the NP within the RC (subject in ORCs, object in SRCs). In this case, gender was always mismatched between the target and distractor items. I consider number to be a *relevant* feature in this experiment, as it is the only feature required for subject-verb agreement in Portuguese. The second self-paced reading experiment was designed particularly to address the question of irrelevant features on retrieval. In these experiments, gender

was manipulated between the target and distractor, but the gender feature is not required to resolve the subject-verb dependency. Across these items, number was matched, and the reasoning for this manipulation will be described in section 3.7 and related sub-sections.

Unlike the first two experiments, a third experiment was conducted to address the question of cue-additivity when a relevant and irrelevant cue are combined, as well as the encoding hypothesis. A Latin square design was also used in this experiment, but instead of crossing feature and clause type, this experiment only contained ORCs and crossed Number (Match, Mismatch) and Gender (Match, Mismatch). This design was used because similarity-based interference effects should primarily occur in ORCs, therefore using only ORCs increases the likelihood that both number and gender would cause processing delays, without diluting the power of the experiment by adding another two-level factor.

3.5.6 General Analysis

Analyses were conducted in R (R Core Team, 2018) using the lme4 package (Bates et al., 2015). In all of the experiments, three separate models were considered. The baseline model contained Clause Type * Feature Match for Experiments 1 and 2 and Gender Match * Number Match for Experiment 3, as well as random intercepts of Subject and Item. The baseline model was not the best fit in any of the experiments reported below when compared to more complex models that considered Trial effects. Trial effects were included in the model because the participants at this university were not accustomed to participating in experiments, and I predicted that they would experience experiment-related fatigue and/or habituation. Goodness of fit was determined by pairwise chi-squared tests using the anova() function in R, and the comparisons are described below

After determining that a model containing Trial was a better fit for the data overall, I then compared whether treating trial as an interactive or an additive predictor was a better fit. This was done by generating two separate models: (1) Match * Clause Type * Trial + (1 | Subject) + (1 | Item) and (2) Match * Clause Type + Trial + (1 | Subject) + (1 | Item). In all experiments, the trial models were first compared against the baseline model. The anova tested had the following format: [Baseline * Trial (Interactive)] and [Baseline * Trial (Additive)] and both models were a better fit for the data across all experiments than the Baseline model (p's <0.001, respectively). From this point, I conducted a third pairwise chi-squared test between models [Trial (Interactive)] and [Trial (Additive)]. and selected the best fitting model in each experiment. The experiments varied on which of the two Trial models best fit the data, but this will be explained in the analysis section of the individual experiments.

The data in these experiments were treated for outliers using the Winsorization method (Dixon, 1960; Tukey, 1962). Unlike outlier removal methods that completely eliminate extreme outliers at a specified threshold a (e.g. all data +/- 3 standard deviations away from the grand mean), winsorization transforms the top and bottom 5% of data to the upper and lower 95th percentile, respectively. This method avoids data loss while also maintaining symmetry in the outlier removal process. This method was selected because the data sets were relatively small, and "elimination-based" outlier methods resulted in up to 11% data loss in certain experimental conditions.[9]

[9] All experiments were also treated for outliers using the 1.5 X IQR removal method. This method eliminates outliers that are 1.5 * the Interquartile Range above the 3rd quartile and below the 1st quartile. Data loss per condition ranged between 6-11%.

3.6 Experiment 1: Number Relative Clauses

Experiment 1 crossed Number (Match, Mismatch) by Clause Type (ORC, SRC). In this experiment, I expected to see the strongest effect of similarity-based interference. As previously mentioned, number is the only relevant feature in resolving subject-verb dependencies in Portuguese. Gender was mismatched across all NPs to prevent any additive effects of interference. This experiment was designed primarily to verify that number that a relevant retrieval cue and capable of generating interference effects in Brazilian Portuguese. Additionally, this experiment cannot directly distinguish between different accounts of the RC asymmetry described in the introduction, but I take evidence showing a processing penalty in object-extracted RCs that contain a number match as support of a standard cue-based retrieval model. This is where retrieval of the subject must occur over the intervening RC subject, which will share number features *and* nominative case.

3.6.1 Design and Materials

As mentioned above, this experiment employed a Latin Square design crossing Number (Match, Mismatch) and Clause Type (SRC, ORC). This design, when combined with grammatical requirements of Portuguese, created slight variations in the verb features across the experimental conditions. In particular, in the case of number-mismatched ORCs, in which the matrix subject and RC subject did not share the same number feature, the matrix and RC verbs have different number morphology to agree with their respective subject (see sample item below).

SRC - Number Match	O homem	que	viu	a menina	na segunda-feira	dá	aula de violão no centro da cidade.
	The man.M.SG.	that	saw.SG.	the girl-F.SG.	on Monday	teaches-SG.	guitar classes in the city center.
SRC - Number Mismatch	O homem	que	viu	as meninas	na segunda-feira	dá	aula de violão no centro da cidade
	The man.M.SG.	that	saw.SG.	the girl.F.PL.	on Monday	teaches.SG.	guitar classes in the city center.
ORC - Number Match	O homem	que	a menina	viu	na segunda-feira	dá	aula de violão no centro da cidade.
	The man.M.SG.	that	the girl.F.SG.	saw.SG.	on Monday	teaches.SG.	guitar classes in the city center.
ORC - Number Mismatch	O homem	que	as meninas	**viram**	na segunda-feira	dá	aula de violão no centro da cidade.
	The man.M.SG.	that	the girls.F.PL.	saw.PL.	on Monday	teaches.SG.	guitar classes in the city center.

Table 2*: Sample item for Number Relative Clause Experiment*

Since this variation is inevitable, the study proceeded with this design. Crucially, the ORC Number Mismatch condition was not predicted to have interference effects, as the number feature of the distractor does not overlap with the target. However, this grammatical variation could potentially make the number mismatch ORC easier to processes than the Match case for non-interference related reasons, as one ORC condition contains interference and the other is completely disambiguated before arriving at the matrix verb. It is unclear if this feature would make the ORC number mismatch case behave more similarly to SRCs in Portuguese when compared to languages such as English where number is less overtly marked on the verb. This concept will be revisited in the general discussion of all experiments.

3.6.2 Analysis

Five sentential regions were analyzed in this study: the relative clause region contained the entire relative clause, the relative clause spillover region contained an unambiguous PP, the matrix verb and object, a spillover region and sentence final region. The baselines for the LMER models were set using contrast coding. SRCs were the baseline for the clause type manipulation and Number Mismatch was the baseline for the feature manipulation, as there were predicted to be theoretically

simplest conditions. As described in the general analysis, I compared three models: the baseline (Number Match * Clause Type), Trial as an additive predictor (Number Match * Clause Type + Trial), and), Trial as an interactive predictor (Number Match * Clause Type*Trial), using a series of pairwise anova tests. In all regions, the models containing Trial were a better fit than the baseline model. However, there was no significant difference in model fit between the two Trial models in most regions. Region 3 showed that the model with Trial and an interactive predictor was a better fit ($X^2(11)=9.78, p<0.05$), but since the simpler trial model (additive predictor) was able to account for the majority of the regions, I will be reporting results from this model.

3.6.3 Predictions

To briefly recap the hypotheses and predictions stated in section 3.5.1, this experiment was designed to test the relevancy of number as a retrieval cue in the resolution of RCs in Brazilian Portuguese. Since number is a relevant retrieval cue, similar to the semantic cues examined in Gordon et al., (among others), it was predicted that Number Match should interact with Clause Type, making Number matched ORCs the hardest to process. In addition, there should be no main effect of Number Match in this experiment, since only ORCs are predicted to generate strong similarity-based interference effects. Additionally, I predict a main effect showing that ORCs are harder to process than SRCs, as has been shown in several processing studies. A main effect of ORC, in conjunction with no main effect of number, also generates an interesting discussion about the nature of relevant cues in retrieval. If we observe this pattern, it would appear that the [+Nominative] match between the matrix and embedded subjects is the strongest source of similarity-based interference, thereby suggesting that a hierarchy exists between retrieval cues. This specific finding may suggest that the retrieval mechanism is initially constrained by syntactic

information and that any other feature matches, such as morphological information, are secondary to syntactic information (see discussion on cue constraints in Chapter 1).

3.6.4 Results

The results for this experiment will be presented by effect, as opposed by region; however, a by-regions graph and the means by region are presented below for reference. All LMER results can be found in Table 3.

	Relative Clause		
	Match	**Mismatch**	**Match Penalty**
ORC	1567 (50)	1520 (53)	47
SRC	1452 (45)	1509 (55)	-57
ORC Penalty	115	11	
		RC Spillover	
	Match	**Mismatch**	**Match Penalty**
ORC	1062 (28)	987 (29)	75
SRC	991 (31)	1007 (30)	-16
ORC Penalty	71	-20	
		Matrix Verb and Object	
	Match	**Mismatch**	**Match Penalty**
ORC	972 (28)	987 (26)	-15
SRC	987 (28)	1005 (30)	-18
ORC Penalty	-15	-18	
		Spillover	
	Match	**Mismatch**	**Match Penalty**
ORC	753 (20)	757 (19)	-4
SRC	775 (19)	770 (21)	5
ORC Penalty	-22	-13	
		Final	
	Match	**Mismatch**	**Match Penalty**
ORC	713 (19)	737 (25)	-24
SRC	767 (24)	731 (21)	36
ORC Penalty	-54	6	

Table 3: *Means and standard errors for reading times in ms (Number RC Experiment)*

	Estimate	Std. Error	df	t value	Pr(>|t|)
RC					
(Intercept)	1845.630	78.512	71.273	23.508	<0.001
TypeObject	43.559	20.417	585.097	2.133	<0.05
MatchingMatch	5.823	20.392	585.059	0.286	0.775
Trial	-31.765	3.624	597.496	-8.765	<0.001
TypeObject:MatchingMatch	30.171	20.375	585.033	1.481	0.139
RC Spillover					
(Intercept)	1164.71	52.303	60.253	22.268	<0.001
TypeObject	18.296	13.306	585.056	1.375	0.170
MatchingMatch	18.708	13.289	585.033	1.408	0.160
Trial	-14.581	2.368	593.103	-6.157	<0.001
TypeObject:MatchingMatch	24.561	13.278	585.018	1.850	0.065
Matrix Verb					
(Intercept)	1147.593	42.644	66.430	26.911	<0.001
TypeObject	-2.614	11.785	585.063	-0.222	0.825
MatchingMatch	-4.428	11.770	585.035	-0.376	0.707
Trial	-15.209	2.095	594.926	-7.258	<0.001
TypeObject:MatchingMatch	2.636	11.760	585.015	0.224	0.823
Matrix Verb Spillover					
(Intercept)	849.273	35.630	54.778	23.836	<0.001
TypeObject	-5.827	7.504	585.030	-0.777	0.438
MatchingMatch	2.435	7.495	585.015	0.325	0.745
Trial	-8.125	1.337	590.593	-6.075	<0.001
TypeObject:MatchingMatch	-1.161	7.489	585.005	-0.155	0.877
Final Region					
(Intercept)	859.183	37.920	64.712	22.658	<0.001
TypeObject	-7.737	8.628	584.961	-0.897	0.370
MatchingMatch	5.985	8.617	584.938	0.694	0.488
Trial	-11.637	1.535	593.320	-7.579	<0.001
TypeObject:MatchingMatch	-13.440	8.610	584.922	-1.561	0.119

Table 4: *LMER Results - Number RCs*

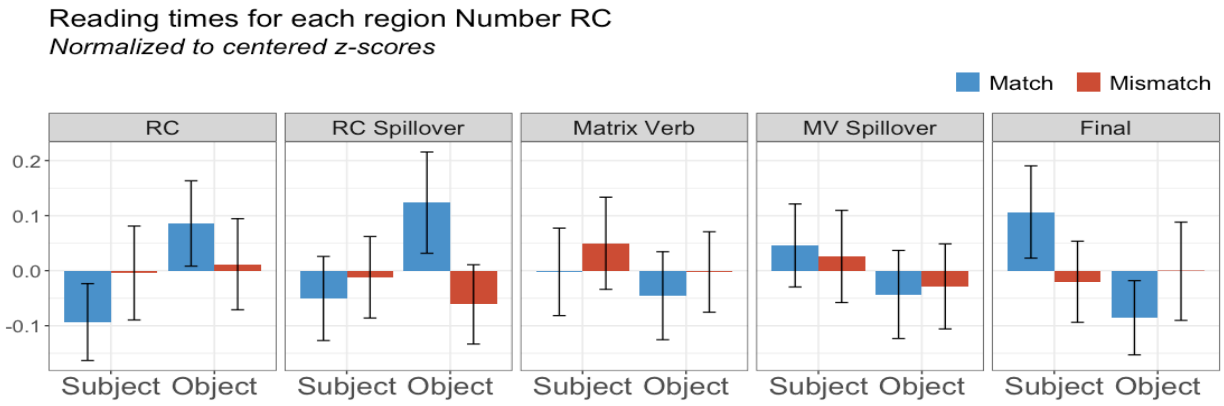


Figure 5: *Graph of normalized to center z-scores for reading times: Number RC Experiment*

3.6.4.1 Object Extraction Penalty

As previous studies have demonstrated, ORCs tend to be harder to process than SRCs. To the best of our knowledge, only one previous study has found this effect in Brazilian Portuguese (Gouvea, 2003). The ORC penalty was observed in the means of the relative clause region (126 ms) and the relative clause spillover region (51 ms) but was only significant in the RC region itself ($p<0.01$). Although we had expected the ORC penalty to be persistent throughout the sentence, the fact that it showed up at the RC region itself is sufficient evidence to show that Brazilian Portuguese speakers are indeed sensitive to the RC asymmetry. More related to the hypotheses, if we were to assume that interference plays a role in the RC asymmetry, it does appear that from a purely cue-based approach, [+Nominative] overlap does generate similarity-based interference between these two structures.

3.6.4.2 Number Match Penalty

There was no significant effect of Number Match in any region. Since SRCs should not contain number interference, I did not predict an across-the-board penalty for Number Match. These findings cannot differentiate between any of the retrieval hypotheses, but they do support the idea

that number matches, although relevant in subject-verb dependencies, should only appear in conjunction with clause type information.

3.6.4.3 Trial

There was a main effect of Trial across all regions ($p<0.01$) indicating that participants read faster towards the end of the experiment. These findings are common in experiments, as subjects become more familiar with the experimental method and perhaps become fatigued with the task. These findings are important to mention because they help to explain why the ORC effect was not present in all of the analyzed regions. If subjects began to read faster or process the sentences in a shallower manner as they progressed in the experiment, we have a potential explanation as to why the ORC penalty was not as pervasive as initially expected. One thought is that subjects could have developed a strategy to focus primarily on the RC region, which contained the crucial manipulation, and pay less attention to the following regions. If this strategy was strengthened throughout the experimental learning that subjects experience, it could have contributed to the containment of our predicted effects to the RC region itself. It is hard to tell, however, if that is the case because subjects were still very successful at responding to the comprehension questions.

3.6.4.4 ORC and Number Match Interaction

In this study, I expected to find an interaction between Clause Type and Number Match demonstrating a reading time penalty for ORCs containing a Number Match between the target and distractor NPs. An interaction would indicate that structural information influenced the retrieval mechanism's selection of the antecedent, since number was not strong enough to cause interference on its own. Although retrieval occurs in both SRCs and ORCs, ORCs force the retrieval mechanism to retrieve a target over an intervening distractor that shares subjecthood features [+Nominative] with the matrix subject, as well as other potential features (e.g. Number in

this manipulation). The increased quantity of overlapping features between the target and distractor in ORCs contributes to the greater amount of similarity-based interference found in these constructions.

A trend for the predicted interaction between Clause Type and Number Match was observed in the post-RC region ($p=0.06$). The weakness and location of this interaction were unexpected given the predictions made by all of the retrieval hypotheses and the overall relevancy of number in subject-verb dependency resolution in Portuguese. The initial prediction was that this interaction would surface at the site of retrieval matrix verb), but this occurred immediately after the RC region. This is unexpected but could be driven by the fact that Portuguese RCs contain more morphological information than English RCs, whereby allowing the comprehender to have sufficient grammatical information available to resolve the embedded clause *before* the matrix verb to resolve the dependency. Since the embedded clause can be fully resolved before the matrix verb, due to the morphological information available, perhaps the embedded subject is no longer a viable distractor in Portuguese RCs (since it's grammatical function in the RC is fully specified).

3.6.5 Comprehension Questions

In addition to measuring reading times, subjects were also presented with a comprehension question after half of the relative clause items (a total of 10 questions). The questions were about the context in the sentence and did not ask about the interpretation of the relative clause itself. The predictions for offline effects reflect those that were made for the online effects: a main effect of Clause Type on comprehension question accuracy and an interaction between Clause Type and Number Match. Overall, there is no support for any effect of either factor on question accuracy in a logistic mixed effects regression model.

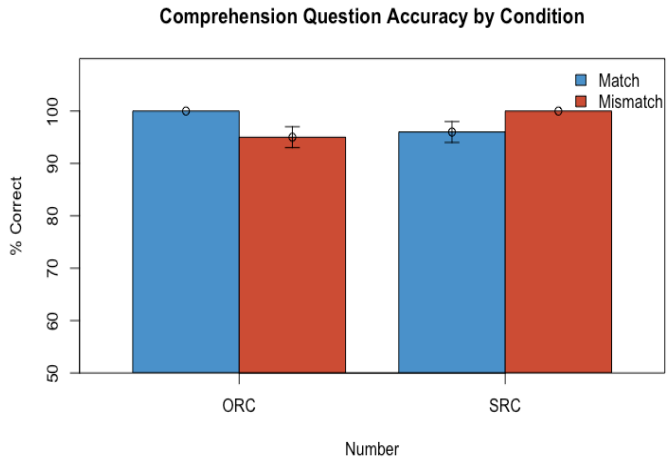


Figure 6: *Graph of Comprehension Question Accuracy Means (Number RC)*

	ORC	**SRC**	**ORC Penalty**
Match	100% (0)	96% (2)	4%
Mismatch	95% (2)	100% (0)	-5%
Match Penalty	5%	-4%	

Table 5: *Comprehension Question Accuracy (Means and Standard Errors) - Number RC*

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	15.240	953.643	0.016	0.987
MatchingMatch	-1.136	559.996	-0.002	0.998
TypeObject	-1.306	559.996	-0.002	0.998
MatchingMatch:TypeObject	10.798	953.645	0.011	0.991

Table 6: *LMER Results - Number RC Questions*

3.6.6 Discussion

The primary predictions for this experiment were to find a penalty for ORCs, as well as an interaction between Clause Type and Number Match. These predictions were inspired by the previous literature on the role of retrieval in RC (Gordon, 2001) and align with classic models of cue-based retrieval in which a linguistic cue can modulate the degree of processing difficulty associated with certain constructions. The findings in this experiment do support these initial

predictions, although not exactly as expected. First off, the ORC penalty appeared only on the relative clause itself and was resolved extremely quickly. In fact, the ORC penalty was resolved so rapidly that there was actually a numerical trend showing an advantage for ORCs after the matrix verb region. Secondly, I did not find the interaction between Number Match and ORC at the retrieval site itself (the matrix verb). Instead, a trend for this interaction appeared on the RC spillover region. The location of this effect may have been a result of the experimental paradigm, as SPR methods tend to delay effects, so this finding can likely be interpreted as difficulty on the RC region itself that carried over into the remainder of the RC (the spillover), but further evidence would be needed to support this claim. It would be interesting to investigate using another segmentation style (such as word-by-word SPR), but given the design of this experiment, these results are not particularly concerning.

Overall, these findings align with the Relevant Feature Hypothesis, stating that linguistic features relevant to the resolution of a grammaticality dependency do indeed affect the retrieval process and generate similarity-based interference effects. In this experiment, number is the only morphosyntactic feature required to resolve the subject-verb dependency present in relative clauses and is strong enough to modulate the RC asymmetry, as demonstrated by the marginal interaction found in the post-RC region. However, since there is no manipulation of the irrelevant feature of gender, this experiment alone cannot tease apart the various retrieval hypotheses.

The potential theoretical contributions of this experiment are (1) number serves as a retrieval cue in ORCs, (2) the strength of morphological feature interference is affected by the overlap of stronger syntactic features (Case, in ORCs) (3) the RC asymmetry is indeed modulated by the degree of overlap between retrieval cues and their potential target + competitors. Expanding

beyond Brazilian Portuguese, these findings should be replicable in all languages that require number to resolve subject-verb dependencies in ORC constructions.

3.7 Experiment 2: Gender Relative Clauses

Unlike Experiment 1, this experiment was designed to test the behavior of grammatical gender in retrieval when it is an irrelevant feature. Since Portuguese grammar expresses only number on the verb, which then projects a retrieval cue seeking an agreeing subject, gender does not participate in agreement procedures (see 17). However, Portuguese contains DP internal gender agreement, which means that speakers of this language must attend to this information at some level of processing in order to extract the meaning of the sentence.

(17) A menina que estuda muito finalmente passou no vestibular da Unicamp.
The $_{F.SG.}$ girl $_{F.SG.}$ that studies $_{SG.}$ a lot finally passed $_{SG.}$ the entrance exam for Unicamp.

The primary motivation behind testing gender morphology was to explore hypotheses about the effects of non-retrieval features on retrieval and/or encoding processes, but this was challenging to achieve since both number and gender are always represented on Portuguese nouns. Portuguese nouns, at least those that follow standard morphological patterns, mark both gender and number in the following format: Root + Gender (*-o*-$_{MAS}$ or *-a*-$_{FEM}$) + Number (*-s*$_{PL}$ or Ø-$_{SG}$), which can be observed in Table 7. For this reason, gender and number features cannot be separated from one another.

	Masculine	**Feminine**
Singular	Gat-o	Gat-a
Plural	Gat-os	Gat-as

Table 7: *Standard Noun Morphology in Portuguese*

In Experiment 1, I made sure that gender cues did not match between the target and distractor, in order to minimize potential gender interference effects. This generated a potential design flaw, as discussed in section 3.6.1, as this manipulation caused one condition to express different verb

morphology on the RC verb than all of the other experiments. To control for this subtle difference between conditions, I decided to match the number feature between the target and distractor so that the verbal morphology was identical across conditions in Experiment 2.

As a result of this manipulation, it is possible that a baseline of number interference already exists for all of the experimental items. Therefore, any effect of gender shown in this experiment cannot be considered to be purely gender interference, as it could also be an additive interference effect. This is important, as the effects observed in this study cannot be used to fully disentangle the predictions made by the Cue (Super)Additivity and Encoding hypotheses, but gender effects *would* provide evidence against a strictly Relevant Feature only hypothesis. With that being said, Experiment 3 was designed to directly address the other hypotheses and will be discussed in section 3.8.

3.7.1 Design and Materials

As in Experiment 1, this experiment used a 2X2 design crossing Gender (Match, Mismatch) X Clause Type (SRC, ORC). As previously mentioned, number features were matched between the target and the distractor. The sample item below shows how gender features and clause type were distributed across conditions.

SRC **Gender Match**	O homem que viu o menino na segunda-feira dá aula de violão no centro da cidade. The man-M.SG. who saw-SG. the boy-M.SG on Monday teaches-SG. guitar classes in the city center.
SRC **Gender Mismatch**	O homem que viu a menina na segunda-feira dá aula de violão no centro da cidade. The man-M.SG who saw-SG. the girl-FEM.SG on Monday teaches-SG. guitar classes in the city center.
ORC **Gender Match**	O homem que o menino viu na segunda-feira dá aula de violão no centro da cidade. The man-M.SG who the boy-M.SG saw-SG. on Monday teaches-SG. guitar classes in the city center.
ORC **Gender Mismatch**	O homem que a menina viu na segunda-feira dá aula de violão no centro da cidade. The man-M.SG who the girl -FEM.SG saw-SG. on Monday teaches-SG. guitar classes in the city center.

Table 8: *Sample item for Gender Relative Clause Experiment*

3.7.2 Analysis

All analyses were conducted using the lme4 package on R. The model of best fit was determined by using paired anova tests between three theoretically motivated models, all containing random intercepts for subject and item (see general analysis for discussion of models). In all comparisons, the models containing Trial were a better fit than the baseline for the data (*p's*<0.05). In the comparison between the Trial models (additive vs. interactive predictor), the results for best fit were mixed. In regions 1, 2, and 5, the model treating Trial as an interactive predictor is the best fit (*p's*<0.05), but there is no significant difference between the models in Regions 3 and 4 (*p's*>0.05). Since the interactive predictor model was a significantly better fit for the majority of the reasons, I report the results from this model below.

3.7.3 Predictions

Given that gender is grammatically required by the language but not overtly involved in subject-verb dependencies, the results of this experiment can help to determine which of the theoretical hypotheses presented in the introduction has the most explanatory power. Under the strongest version of the Relevant Feature Hypothesis, Gender should not have any effect on relative clause processing in Portuguese. This leads to the prediction that Gender Match should not interact with

Clause Type, as was predicted for Number Matches in Experiment 1. However, if the interaction is present, it would be difficult to determine whether the Relevant Feature Hypothesis is incorrect or if the interactions support the Cue (Super)Additivity Hypothesis. As it is formulated, the Cue (Super)Additivity Hypothesis would also predict an interaction between Gender Match and Clause Type the current experiment, because there is a baseline of relevant feature interference (Number) included in the design that could be strengthened by matching non-retrieval features (Gender). It is important to note, however, that this experiment was not designed to definitively separate these two hypotheses, but rather to set the stage for Experiment 3.

The absence of an interaction can also be informative. If no interaction is observed in the experiment but there is a penalty for Gender Match, there would be strong support for the Encoding Hypothesis. Since Gender is not required to resolve subject-verb dependencies in Portuguese, a penalty for Gender Match cannot be attributed to retrieval interference and might be caused by an alternative process affecting sentence processing. As mentioned in the introductory chapter, a very similar study on Italian relative clauses found what appears to be encoding interference produced by gender matches (Villata et al. 2018). Given that gender in this experiment should not provoke retrieval interference, I adopt the viewpoint of these authors that any slowdown caused by gender may be attributed to encoding processes.

3.7.4 Results

As in Experiment 1, the results are presented by effect and their corresponding predictions. The normalized means used in the analysis are presented in the table below and the corresponding LMER results are in Table 9.

	Relative Clause		
	Match	**Mismatch**	**Match Penalty**
ORC	1922 (83)	1964 (74)	-42
SRC	1820 (71)	1708 (62)	112
ORC Penalty	102	256	
	RC Spillover		
	Match	**Mismatch**	**Match Penalty**
ORC	1010 (31)	1125 (40)	-115
SRC	1120 (37)	1051 (25)	69
ORC Penalty	-110	74	
	Matrix Verb and Object		
	Match	**Mismatch**	**Match Penalty**
ORC	994 (33)	1067 (34)	-73
SRC	1093 (36)	1101 (35)	-8
ORC Penalty	-99	-34	
	Spillover		
	Match	**Mismatch**	**Match Penalty**
ORC	849 (30)	815 (23)	34
SRC	820 (24)	805 (23)	15
ORC Penalty	29	10	
	Final		
	Match	**Mismatch**	**Match Penalty**
ORC	678 (23)	654 (18)	24
SRC	721 (22)	751 (25)	-30
ORC Penalty	-43	-97	

Table 9: *Means and standard errors for reading times in ms (Gender RC Experiment)*

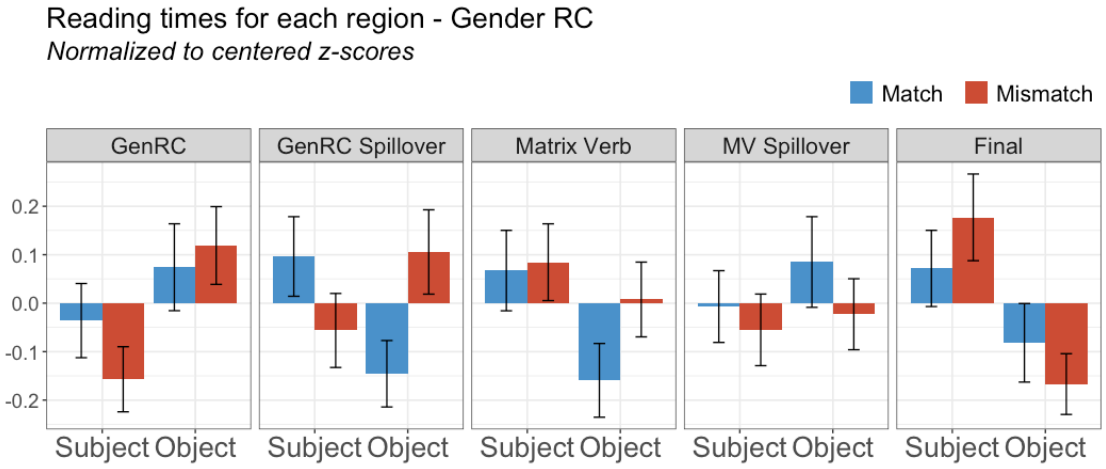


Figure 7: *Graph showing normalized to center z-scores for reading times: Gender RC Experiment*

	Estimate	Std. Error	df	t value	Pr(>|t|)
RC					
(Intercept)	2275.119	119.264	67.078	19.076	<0.001
TypeObject	157.657	61.089	589.631	2.581	<0.05
MatchingMatch	189.150	61.384	590.831	3.081	<0.01
Trial	-39.834	5.037	586.281	-7.908	<0.001
TypeObject:MatchingMatch	10.433	61.475	591.837	0.170	0.865
TypeObject:Trial	-3.974	5.122	591.512	-0.776	0.438
MatchingMatch:Trial	-15.960	5.153	593.014	-3.097	<0.01
TypeObject:MatchingMatch:Trial	-3.374	5.163	594.258	-0.653	0.514
RC Spillover					
(Intercept)	1274.642	55.089	66.901	23.138	<0.001
TypeObject	-25.462	31.888	592.168	-0.798	0.425
MatchingMatch	-10.402	32.022	594.554	-0.325	0.745
Trial	-19.090	2.632	586.467	-7.253	<0.001
TypeObject:MatchingMatch	-123.778	32.077	595.153	-3.859	<0.001
TypeObject:Trial	2.688	2.673	594.690	1.006	0.315
MatchingMatch:Trial	-0.488	2.687	597.579	-0.181	0.856
TypeObject:MatchingMatch:Trial	7.573	2.693	598.340	2.812	<0.01
Matrix Verb					
(Intercept)	1246.963	55.342	69.554	22.532	<0.001
TypeObject	-57.965	29.936	590.633	-1.936	0.053
MatchingMatch	-16.362	30.078	592.006	-0.544	0.587
Trial	-17.647	2.469	586.809	-7.148	<0.001
TypeObject:MatchingMatch	-12.256	30.121	593.104	-0.407	0.684
TypeObject:Trial	3.413	2.510	592.744	1.360	0.174
MatchingMatch:Trial	-0.316	2.525	594.459	-0.125	0.900
TypeObject:MatchingMatch:Trial	-0.261	2.530	595.809	-0.103	0.918
Matrix Verb Spillover					
(Intercept)	969.681	40.633	67.092	23.865	<0.001
TypeObject	-3.983	21.013	589.764	-0.190	0.850
MatchingMatch	24.683	21.113	591.084	1.169	0.243
Trial	-14.118	1.733	586.214	-8.148	<0.001
TypeObject:MatchingMatch	40.850	21.145	592.014	1.932	0.054
TypeObject:Trial	2.189	1.762	591.678	1.243	0.215
MatchingMatch:Trial	-0.941	1.772	593.323	-0.531	0.596
TypeObject:MatchingMatch:Trial	-3.285	1.776	594.475	-1.850	= 0.0648
Final Region					
(Intercept)	855.897	35.476	66.665	24.126	<0.001
TypeObject	-63.181	17.831	589.589	-3.543	<0.001
MatchingMatch	20.631	17.919	590.548	1.151	0.250
Trial	-14.948	1.470	586.617	-10.169	<0.001
TypeObject:MatchingMatch	35.668	17.944	591.721	1.988	<0.05
TypeObject:Trial	3.604	1.495	591.418	2.411	<0.05
MatchingMatch:Trial	-1.960	1.504	592.635	-1.303	0.193
TypeObject:MatchingMatch:Trial	-1.901	1.507	594.079	-1.262	0.208

Table 10: *LMER Results for Gender RCs*

3.7.4.1 ORC Penalty

A trend for an ORC penalty was found in the RC region (358 ms penalty, $p<0.05$). These results align with the prediction that ORCs are harder to process than SRCs. It is also important to note that, as observed in Experiment 1, the ORC penalty was quickly resolved and only appeared in the RC region. Although this could be an experimental artifact, these findings should also be explored in other morphologically rich languages to see if this is a result of having more grammatical information available within the RC allows for faster processing (e.g. overtly marked verb morphology, indicated below by a '**ø**'), as opposed to having to wait until the matrix verb for unambiguous dependency resolution (see example 16). Crucially, the nearly identical results between both experiments strongly suggest that ORCs incur reading time penalties regardless of the type of cue being manipulated in Brazilian Portuguese.

(16) a. The friends who Mary met in Italy visited Los Angeles in August.
The friends$_{PL.}$ who Mary$_{SG.}$ **met ø** in Italy **visited ø** Los Angeles in August.

b. Os amigos que Maria conheceu na Italia visitaram Los Angeles em agosto.
The friends$_{PL.}$ who Mary$_{SG.}$ **met SG.** in Italy **visited PL.** Los Angeles in August.

Further evidence that the ORC penalty is quickly resolved in Portuguese is shown by a flip towards a marginal ORC advantage at the matrix verb region ($p=0.05$), which becomes significant in the final region ($p<0.001$). Experiment 1 also showed a numerical trend for this reversal towards the end of the sentence. The repetition of this effect raises the question of why these participants seemed to only struggle with ORCs at the RC region itself. As discussed above, this could be a grammatical effect because Portuguese speakers have more agreement information available to them at the RC verb, but this is unlikely to be the only explanation. In (17a), this logic could be applied, as the RC internal subject-verb agreement eliminates the option for the RC subject to be

a potential target for the matrix verb. However, the items in this experiment were all matched on number features, so the RC subject could still be a potential candidate (17b).

(17) a. Os pacientes que o médico tratou já foram para casa.
The$_{M.PL.}$ patients$_{M.PL.}$ that the$_{M.SG.}$ doctor$_{M.SG.}$ treated$_{SG.}$ already went$_{PL.}$ home.

b. Os pacientes que os médicos trataram já foram para casa.
The$_{M.PL.}$ patients$_{M.PL.}$ that the$_{M.PL.}$ doctors$_{M.PL.}$ treated$_{PL.}$ already went$_{PL.}$ home.

Since this experiment followed the format of (17b), a purely grammatical account cannot explain the rapid resolution of ORCs, nor the preference towards SRCs at the end of the sentence. This interpretation should not be completely eliminated because it could still account for cross-linguistic differences in RC processing (see 16), but it is that the ORC advantage effects are some sort of experimental artifact, such as habituation or segmentation of the sentences. In order to determine which of these is the cause, a word-by-word SPR study could be conducted to see if segmentation matters or habituation effects could be looked for in items included in this overall design. For now, I take the ORC penalty at the RC region itself to be the most informative and theoretically relevant finding in relation to our initial predictions and reserve the discussion of the reversal of this pattern for future research.

3.7.4.2 Gender Match Penalty

This primary goal of this study was to determine if gender features cause any sort of processing disruption, as this would warrant speculation about the retrieval mechanism or be indicative of a secondary process affecting the processing of RCs, such as encoding interference. There was a Gender Match penalty in the relative clause region (70 ms penalty, $p<0.01$). These findings are quite interesting from the theoretical perspective because none of the retrieval-driven hypotheses would predict that a main effect of gender would appear on its own. Even under the two hypotheses where irrelevant features are considered, we have no evidence that *relevant* features produce main

effects, since number did not produce a main effect in Experiment 1, whereby demonstrating that the effects are neither identical nor super additive.

Moreover, these effects appeared before the retrieval site. Although it was not possible with the SPR segmentation to see the exact location that these effects appeared, the fact that they showed up before the Matrix Verb aligns closely with the encoding hypothesis (Van Dyke and McElree, 2006 for discussion) As discussed in the introduction, any interference caused by these retrieval-irrelevant features may support claims that the encoding process also produces strong enough similarity-based interference effects to directly impact sentence processing.

3.7.4.3 Trial

Subjects read the sentences faster as the experiment progressed, as evidenced by the Trial effects present in all regions (p's<0.001 in all analyzed regions). Trial also interacted with the factors of Match and Clause Type throughout the sentence. In the RC region, sentences containing Gender Match were read faster as subjects progressed through the experiment (p<0.001). There was also a marginal three-way interaction showing that Gender Matched ORCs were read faster throughout the experiment in the matrix verb spillover region (p=0.06). This is reminiscent of Experiment 1, which also showed that the Number Match - ORC condition was read faster over time.

What is puzzling, however, is that the opposite three-way interaction between Clause Type, Gender Match and Trial appeared in the RC spillover region – showing that this region gets harder over time (p<0.01). This implies that subjects started to struggle with the post-RC region as the experiment progressed, which seemingly contradicts the finding in the paragraph above showing that Gender Match - ORCs got easier throughout the experiment. A potential explanation for this could be a clause-level wrap-up effect. Essentially, subjects could have developed a strategy throughout the experiment to wait for the presentation of the entire RC before deeply processing

the SRC or ORC distinction. This could cause them to pause briefly on the RC spillover region and then progress to the later sentential regions.

Another difficult-to-interpret interaction is observed in the final region, where ORCs became more challenging as time passed. This interaction contradicts the trend towards ORC facilitation after the matrix verb region found in the raw means and main effects discussed in the previous section. I propose that these effects may be unrelated. Perhaps reading times decreased throughout trials as a byproduct of experimental participation and fatigue, which generated the Trial effects. These faster reading times may have resulted in increasingly shallow sentence processing throughout the experiment, which had to be addressed when subjects had to decide on a final interpretation for the sentence. This interaction is not of great importance for the overall interpretation of the results in Experiment 2, but it raises interesting questions about how effects may change simply as a result of experimental procedures.

3.7.4.4 Interaction between Gender Match and Clause Type

Any form of interaction between Gender Match and Clause Type would be very useful for narrowing down the hypotheses presented earlier in this chapter. For example, the absence of this interaction would provide strong support for the Relevant Feature Retrieval hypothesis, which states that the retrieval mechanism is not sensitive to non-retrieval features, thus aligning with the current literature on cue-based retrieval. On the other hand, the presence of an interaction showing a penalty for Gender Matched ORCs would indicate that non-retrieval features are capable of generating similarity-based interference effects that are strong enough to disrupt sentence processing. I hesitate to make a claim about the source of this interaction, as it is very challenging to determine if the source of this processing disruption is caused by the retrieval or the encoding mechanism. In this experiment, an interaction could result from any of the following mechanisms:

(1) a Generalized Feature Retrieval mechanism, since we saw this effect in Experiment 1 for number, (2) a Cue Additivity process because number (relevant feature) is matched across all conditions, or (3) a combination of relevant feature retrieval and encoding interference occurring at the same time, as several authors have claimed that encoding effects can surface at the retrieval site (see Chapter 1 for discussion).

With these three potential explanations in mind, there are a few small ways that they can be differentiated. First off, we found a main effect of gender in the RC, which was not observed for number in Experiment 1. Since the Generalized Feature Retrieval hypothesis predicts identical behaviors for all cues, this small difference already works against this hypothesis. Separating the Cue Additivity Hypothesis from the Encoding Interference + Relevant Feature Retrieval Hypothesis is slightly harder, but the early main effect of gender aligns more closely with an encoding account than a retrieval account. Regardless of the underlying source of these interference effects. this interaction would generate novel evidence that does not align with the current cue-based retrieval literature. Rather unexpectedly, this experiment showed the presence and absence of this interaction, which is discussed throughout this section.

To begin, the current retrieval literature would lead to the prediction that Gender should not interact with Clause Type at all, as it is not a relevant feature in the subject-verb dependency. In support of this, three of the five analyzed regions showed no interaction between Gender Match and ORC. Moreover, the RC spillover region actually showed a facilitatory interaction for Gender Matched ORCs (46 ms advantage, $p<0.01$). Although a facilitatory interaction was also not expected, this effect is interesting to examine, as it numerically spills over into the Matrix Verb region as well. Looking at the means for the RC spillover and Matrix verb regions, there is a numerical trend showing that ORCs have a reading time advantage and that Gender Matches in

SRCs become increasingly difficult to process (see Figure 5). What this facilitation effect means for the retrieval mechanism is unclear. However, since this dissertation focuses primarily on inhibitory interference, I will not dwell on this topic. Overall, I take the overall absence of an interaction between Clause Type and Gender Match, as well as any sort of facilitation, to mean that non-retrieval features *do not* generate similarity-based interference in sentence processing. Unfortunately, this is not the only interaction observed in the data.

The claims above are challenged at the matrix verb spillover region. In this region, there is an interaction showing that Gender Matched ORCs incurred the slowest reading times. In fact, the trend favoring SRCs and Gender Matches in ORCs from the presentation of the RC through the matrix verb is completely flipped in the matrix verb spillover region ($p<0.001$). Looking at the raw means at the matrix verb spillover region (Table 11), the Gender Match - ORC condition was the slowest by a relatively small amount, but this does not minimize the importance of this finding.

	Match	**Mismatch**	**Match Penalty**
ORC	849 (30)	815 (23)	34
SRC	820 (24)	805 (23)	15
ORC Penalty	29	10	

Table 11: *Means and Standard Errors for Matrix Verb Spillover Region*

Taken at face value, this interaction suggests that the retrieval mechanism is indeed sensitive to *all* feature matches, not just those required to solve the grammatical dependency. However, I think this is relatively unlikely. To begin, the location of this effect immediately discredits any supper additive effect that the Cue Additivity hypothesis would predict. In Experiment 1, a marginal interaction between Number Match (a relevant feature) and Clause Type appeared in the post-RC region ($p=0.06$); therefore, if cue matches were truly super additive, we would have expected this interaction to appear in the same location as it did in Experiment 1, especially since the sentences are identical in both experiments. This line of argumentation also applies to the Generalized

Feature Retrieval Hypothesis, which it predicts that all features should generate identical effects. The combination of the location of this effect and the main effect of gender match found at the RC are not seen when only number is relevant, which demonstrates distinct behaviors for both features.

For these reasons, this interaction most strongly supports a combination of the Relevant Feature Retrieval Hypothesis and the Encoding Hypothesis. I have already discussed in depth why the embedded subject in an ORC causes increased interference effects: the retrieval cue must access the matrix subject over an NP that also has Nominative Case. So, in this sense, the ORC component of the interaction makes sense. We then must address why the encoding effects surfaced in ORCs if they are not directly involved in retrieval. This answer is relatively simple: when a distractor is encoded with an additional feature (gender) that overlaps with the matrix subject, feature overwriting causes the activation levels of both the matrix and embedded NPs to decrease (Villata et al., 2018; see Chapter 1 for detailed description). This overall decrease in activation levels, when combined with the retrieval interference effects that arise in ORCs, allows the encoding effects to surface at the same time. Essentially, encoding interference directly decreased the retrievability of the target before retrieval occurred, which negatively affected the retrieval processes. Therefore, the combination of both encoding and retrieval interference is likely what made Gender and Number Matched ORCs the hardest condition to process.

3.7.5 Comprehension Questions

As in Experiment 1, subjects were also presented with comprehension questions after ½ of the Relative Clauses (10 comprehension questions). The responses to the comprehension questions in this study were of particular interest because Villata et al. (2018) claimed that encoding interference effects are more likely to show up in offline responses than in online studies, based on their experiments and a post-hoc analysis of several others. Unlike the analysis conducted by

Villata et al. on Italian, we see no effect of Gender Match or Clause Type in the comprehension question responses.

	ORC	SRC	ORC Penalty
Match	94% (3)	96% (2)	-2%
Mismatch	96% (2)	95% (2)	1%
Match Penalty	-2%	2%	

Table 12: *% Correct on RC Comprehension Questions – Gender Experiment*

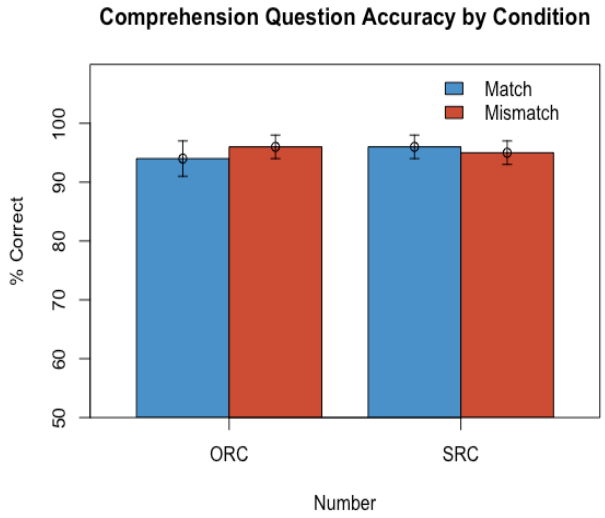


Figure 8: *% Correct on RC Comprehension Questions – Gender Experiment*

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.777	1.120	4.264	<0.001
MatchingMatch	-0.043	0.306	-0.140	0.889
TypeObject	-0.138	0.314	-0.439	0.660
MatchingMatch:TypeObject	-0.239	0.311	-0.766	0.444

Table 13: *GLMER Results - Comprehension Questions in Gender RCs*

3.7.6 Discussion

To summarize the findings, Experiment 2 showed (1) a penalty for Gender Match at the RC region, (2) a penalty for ORCs at the RC region, (3) an interaction between Clause Type and Gender Match showing that Gender Matched ORCs were the hardest sentence to process in the matrix verb

region, and (4) a facilitatory interaction between Clause Type and Gender Match in the RC spillover region. As mentioned in the discussion about the interactions, the exact implications of both the facilitatory and inhibitory interactions are unclear. For now, I will tentatively assume that the facilitatory interaction, supported by the lack of interaction in 3 other sentential regions, means that the retrieval mechanism is not sensitive to irrelevant features. The absence of an interaction demonstrates that irrelevant cues are not able to generate similarity-based interference during retrieval, thereby suggesting that they are not employed by the retrieval mechanism during cue-matching procedures. The inhibitory interaction, on the other hand, appears to be more closely related to the fact that the ORC penalty decreased over time than an actual effect of gender match. Overall, these findings largely support current models of the cue-based retrieval that claim that the retrieval mechanism is content-addressable and has direct access to all information relevant to matching the retrieval cue with its target but also raises questions about the involvement of encoding in sentence processing.

The most theoretically interesting result in this study was the penalty for Gender Match in the RC region. In most of the literature on retrieval, the term *similarity-based interference* is typically associated with a cue-based retrieval mechanism in which any degree of cue-overload makes the [retrieval cue + target] matching procedure less accurate or slower. Crucially, the retrieval cue will only match with its target if the feature relevant to the dependency being resolved. The current results, however, strongly imply that similarity-based interference may not be exclusive to the retrieval process, as gender is not a relevant feature in Portuguese subject-verb dependencies. This claim is supported by the fact that the effect of Gender Match was not limited to environments where retrieval difficulty would be expected (ORCs). For this reason, the gender effects found in this study provide strong with the Encoding Hypothesis, claiming that encoding

interference can indeed result in observable processing penalties. These findings closely align with the studies conducted by Villata et al. (2018) on Italian, who also found a Gender Match effect on ORCs and lend further support to their arguments in favor of encoding interference (see Chapter 2 for a detailed overview of their experiments).

3.8 Experiment 3: Gender + Number ORCs

Experiment 3 was designed to address the behavior of both irrelevant and relevant features more directly. In particular, this experiment addresses the questions brought up by the findings in Experiment 2 showing that Gender Matched ORCs incurred a processing penalty. In particular, whether these results lean more towards a retrieval mechanism that is sensitive to all cues or a highly constrained mechanism that is affected by encoding interference. Relying on these findings of the previous experiment that showed effects for ORC, Gender Match and Number Match, only ORCs were used in this study to more directly observe the effects of morphological feature matches on processing, independent of clause type.

3.8.1 Method and Design

This experiment differed from the previous two because there was no clause type manipulation. Subjects were presented with 20 ORCs that were manipulated for feature matches. The Latin Square design used in this study crossed Number (Match, Mismatch) and Gender (Match, Mismatch) between the target (matrix subject) and subject (RC internal DP). It is also important to note that, as in Experiment 1, the conditions in which number did not match, the matrix verb and RC verb do not possess the same number morphology (in bold in sample item, conditions Number Mismatch, Gender Match/Gender Mismatch). I will return to discussing the potential effects of this in the discussion of the experiment.

Number Match - Gender Match	O homem que o menino viu na segunda-feira dá aula de violão no centro da cidade. The man-M.SG. who the boy-M.SG saw-SG. on Monday teaches -SG. guitar classes in the center of the city.
Number Match - Gender Match	O homem que a menina viu na segunda-feira dá aula de violão no centro da cidade. The man-M.SG who the girl-FEM.SG saw S-G. on Monday teaches-SG. guitar classes in the center of the city.
Number Mismatch - Gender Match	O homem que os meninos **viram** na segunda-feira **dá** aula de violão no centro da cidade. The man-M.SG who the boys-M.PL. saw -PL. on Monday teaches-SG. guitar classes in the center of the city.
Number Mismatch- Gender Mismatch	O homem que as meninas **viram** na segunda-feira **dá** aula de violão no centro da cidade. The man-M.SG who the girls -FEM.PL saw -PL. on Monday teaches-SG. guitar classes in the center of the city.

Table 14: *Sample item for Number X Gender ORCs*

3.8.2 Analysis

All analyses were conducted using the lme4 package in R. As in the previous two experiments, three models were tested, all of which contained random slopes for Item and Subject: (1) a baseline model: Number * Gender, (2) (2) Trial as an additive predictor: Number * Gender + Trial, and (3) Trial as an interactive predictor: Number * Gender * Trial. In all regions, both models containing Trial were a better fit than the baseline (p's<0.05), and in a second pairwise anova test, there was no significant difference in model fit between models treating trial as an additive or interactive predictor (p's>0.05). For this reason, the simpler model (3) was selected and its results are reported below.

3.8.3 Predictions

Considering the findings of Experiments 1 and 2, the findings in this study are extremely important for addressing the hypotheses presented in this chapter. Experiment 1 showed that matching number features modulated the degree of difficulty associated with ORCs, thus supporting the relevant feature hypothesis. Experiment 2, however, showed a main effect of Gender Match, as well penalty between Gender Match – ORC, both of which seem to challenge the predictions of

the Relevant Feature hypothesis, if this were the only mechanism that generates interference. As discussed in Experiment 2, the interaction might have been either super additive retrieval interference or encoding interference, so this experiment will help find the genuine source behind Feature Match – ORC penalties. For only the Relevant Feature hypothesis to be supported, there should be a main effect of Number Match in Experiment 3 but no main effect of Gender Match, as it is not a relevant retrieval cue. It is unlikely that this hypothesis can account for all of the findings though since gender has been shown to produce processing delays in Experiment 2.

This experiment also helps to address the cue-additivity hypothesis, which would generate the prediction that Gender Match and Number Match should interact with each other. Following the hypothesis, when relevant number features match, we should observe a super-additive effect of gender. Essentially, this would mean that the Gender Match – Number Match condition should be the hardest to process. This result is super-additive because it is not predicted under this model that gender should cause any effects on its own. A lack of interaction, following this logic, would provide evidence against the additivity approach and align more closely with the approach that encoding and retrieval interference effects are surfacing in tandem.

Returning to the encoding hypothesis that was supported in Experiment 2, I maintain the same predictions. There is strong evidence that the gender match effects in Experiment 2 are more related to encoding than retrieval due to how early they appear and the fact that gender produces a main effect across clause types. In this experiment, a simple main effect of Gender Match that does not interact with number can also be construed as evidence that Brazilian Portuguese speakers experience a certain level of interference from gender information that cannot only be attributed to retrieval processes.

3.8.4 Results

The findings are presented by predicted effects. I specifically looked for a main effect of Number Match, a main effect of Gender Match and a potential interaction between the two. All of the LMER results can be found Table 15[10].

	Relative Clause		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	1588 (55)	1751(77)	-163
Gender Mismatch	1664 (67)	1477 (49)	187
Gender Match Penalty	-76	274	
	RC Spillover		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	1064 (34)	1127 (41)	-63
Gender Mismatch	1068 (34)	1034 (34)	34
Gender Match Penalty	-4	93	
	Matrix Verb and Object		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	1021 (34)	1046 (36)	-25
Gender Mismatch	984 (1025)	1025 (32)	-41
Gender Match Penalty	37	21	
	Spillover		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	769 (20)	759 (21)	10
Gender Mismatch	754 (21)	752 (19)	2
Gender Match Penalty	15	7	
	Final		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	801 (30)	743 (25)	58
Gender Mismatch	732 (24)	744 (24)	-12
Gender Match Penalty	69	-1	

Table 15*: Means and Standard Errors for Reading Times - Gender X Number ORCs*

[10] A main effect of Trial was found in all regions (p's<0.01)

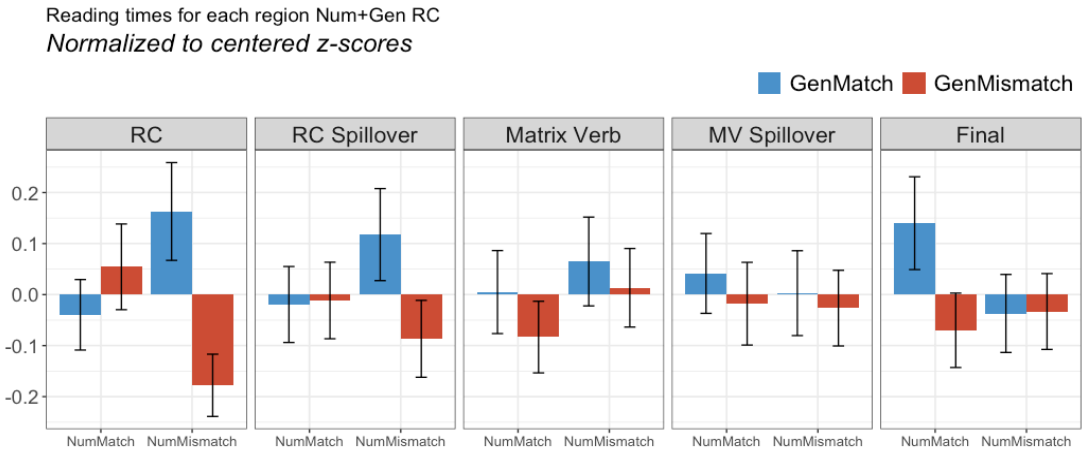


Figure 9: *Graph of Normalized to Centered Z-scores for reading times - Gender X Number ORCS*

	Estimate	Std. Error	df	t value	Pr(>|t|)
RC					
(Intercept)	1951.662	106.857	63.017	18.264	<0.001
genderGenMatch	46.692	23.627	585.009	1.976	<0.05
numberNumMatch	8.723	23.627	585.009	0.369	0.712
Trial	-31.600	4.146	589.859	-7.623	<0.001
genderGenMatch:numberNumMatch	-95.723	23.650	585.019	-4.048	<0.001
RC Spillover					
(Intercept)	1255.358	56.270	63.567	22.309	<0.001
genderGenMatch	20.744	14.924	585.001	1.390	0.165
numberNumMatch	-5.511	14.924	585.001	-0.369	0.712
Trial	-17.362	2.619	589.320	-6.628	<0.001
genderGenMatch:numberNumMatch	-28.718	14.939	585.010	-1.922	0.055
Matrix Verb					
(Intercept)	1177.770	53.754	63.260	21.910	<0.001
genderGenMatch	13.194	13.239	584.985	0.997	0.319
numberNumMatch	-14.994	13.239	584.985	-1.133	0.258
Trial	-15.119	2.324	588.924	-6.506	<0.001
genderGenMatch:numberNumMatch	-0.205	13.252	584.993	-0.015	0.988
Matrix Verb Spillover					
(Intercept)	859.232	32.417	62.347	26.505	<0.001
genderGenMatch	4.820	8.363	585.002	0.576	0.565
numberNumMatch	3.882	8.363	585.002	0.464	0.643
Trial	-9.600	1.468	589.068	-6.540	<0.001
genderGenMatch:numberNumMatch	-0.628	8.371	585.010	-0.075	0.940
Final Region					
(Intercept)	921.827	38.638	72.248	23.858	<0.001
genderGenMatch	15.298	10.345	584.978	1.479	0.140
numberNumMatch	12.760	10.345	584.978	1.233	0.218
Trial	-15.884	1.814	591.403	-8.758	<0.001
genderGenMatch:numberNumMatch	13.141	10.355	584.991	1.269	0.205

Table 16: *LMER Results, Gender + Number ORCs*

3.8.4.1 Number Match

Unlike in Experiment 1 where an interaction between number and clause type was expected, Experiment 3 should have only shown a main effect of Number Match because only ORCs were presented, as no clause type manipulation was included. Unexpectedly, number matches did not result in a reading time penalty in any region, implying that this feature did not generate similarity-based interference effects. As will be discussed in the subsequent sections, the absence of this effect contradicts the predictions of all of the retrieval hypotheses. Although there are several possible reasons why number effects were absent that I will discuss in general discussion, these results automatically cast doubt on the rest of the findings in this study regarding retrieval processes.

3.8.4.2 Gender Match

In the case of Gender Match, there were several possible outcomes that were predicted by each of our initial hypotheses. In the case where only relevant cues cause retrieval interference, there should have been no effect of gender in this experiment. On the other hand, if similarity-based interference effects are generated at the encoding level, as opposed to just being a retrieval effect, a main effect of Gender Match is possible. In Experiment 2, there was indeed a penalty for Gender Match, which provided some support for the hypothesis about encoding interference; however, as mentioned in the introduction to this section, it is impossible to determine if the Gender Match penalty was a result of encoding interference or some sort of additive effect caused by the presence of other interfering material in the sentence. This current experiment was designed largely to provide more conclusive support for either the encoding or additivity hypothesis, as there were conditions where number was mismatched.

With that being said, a penalty for Gender Match occurred in the RC region ($p<0.05$). This is important because, as previously mentioned, we now have evidence that gender causes reading time slowdowns independent of number. Since no super-additive effect was observed, the gender penalty provides more support for my claims regarding encoding interference in Experiment 2, as this effect has now been consistent across two separate studies in the same region. As further evidence, the Gender Match penalty in Experiments 2 and 3 both occur prior to the matrix verb, which is the retrieval site. I mention this cautiously, however, as Experiment 1 showed the predicted retrieval interference effects prior to the matrix verb as well (at the RC-spillover region). Unfortunately, the segmentation of this SPR study does not allow for further time course analyses to be conducted.

3.8.4.3 Interaction between Number and Gender

As in Experiment 2, both the presence or the absence interaction between Gender Match and Number match has the potential to be the most influential finding of this chapter, as it will take us one step closes to narrowing down the original hypotheses and gaining a deeper understanding of the retrieval and encoding mechanisms. Before presenting the results, I will lay out the primary predictions that an interaction would generate. To begin, an interaction showing that Gender Match – Number Match is the hardest condition to process would have two important implications: (1) the retrieval mechanism is sensitive to non-retrieval features and (2) there is an additive effect when more than one feature is matched (Cue (Super)Additivity Hypothesis). Contrastingly, the absence of this interaction would indicate that the retrieval mechanism is not sensitive to non-retrieval features and imply that gender features do not have a strong effect on processing to modulate the RC asymmetry, despite the main effect of gender observed in the RC region. With that being said, the results in this study produce a third type of interaction that was unexpected, a

facilitatory interaction between Gender Match – Number Match in the RC region ($p<0.05$) and in the RC spillover region ($p=0.055$) that I will explore throughout this section

The facilitatory interaction between Gender Match – Number Match is slightly deceiving. The most straightforward interpretation is that when both of these features match, processing is facilitated, perhaps indicating that the retrieval mechanism has become completely overloaded or is failing in some way. Before exploring this option, I would like to break down the interaction more precisely. Looking at the means, there is a 24 ms penalty for Number Match and a 198 ms penalty for Gender Match, which indicates that this interaction is primarily driven by Gender. This is relatively unexpected, as Gender was predicted to have a minimal role in this experiment due to its lack of involvement in the dependency in question.

Digging a bit deeper, a more interesting pattern appears in the means: an alternation in the strength of the penalties for each feature. For example, when Gender is matched, there is *facilitation* for Number Match of 163 ms, but in cases where Gender is mismatched, there is a number match *penalty* of 187 ms. This same pattern surfaces when looking at the effect of Number match/mismatch on Gender effects. In Number Match conditions, there is a 76 ms advantage for Gender Match, whereas, in Number Mismatch conditions, there is a 274 penalty for Gender Match. Importantly, even though this tradeoff is observed – the overall penalty for Gender Match remains

the strongest across conditions. The interaction plot below provides a visual representation of what this tradeoff looks like.

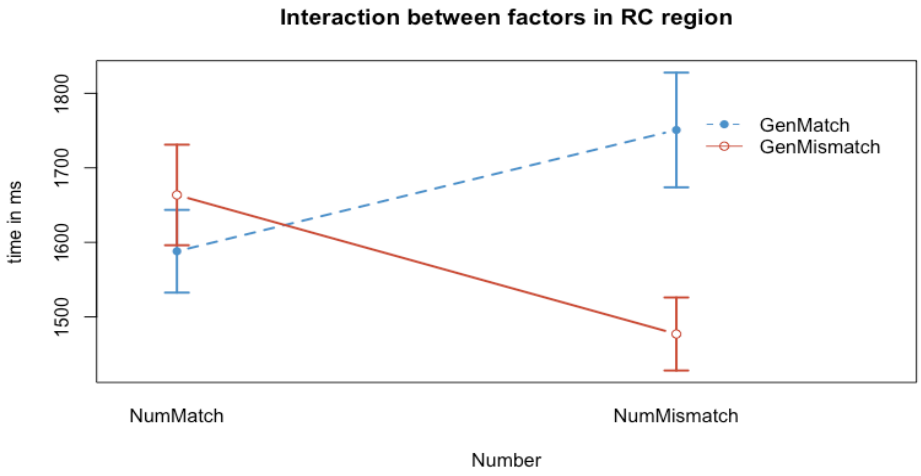


Figure 10: *Interaction Plot Between Gender and Number in RC Region – Gender X Number ORCs*

This tradeoff has interesting implications for the hypotheses related to this experiment. First and foremost, this finding does not support the predictions made by the Cue (Super)Additivity Hypothesis, as there was no observable interaction between Gender Match and Number Match showing that this is the hardest condition to process. Secondly, this finding, when coupled with the main effects of Gender Match in Experiments 2 and 3, provides evidence that non-retrieval cues do indeed generate similarity-based interference effects. Unfortunately, the lack of number match penalties in any region is not compatible with the claims of any hypothesis made in this dissertation. As discussed in the introduction, ORCs should always have a baseline of interference due to the presence of two subjects, which is worsened when number also matches between the target and the distractor. The absence of a number match effect, or even a super-additive interaction between gender and number, raises questions about the strength of number as a retrieval cue in Portuguese.

Stepping away from number, we cannot ignore the surprising behavior of gender in this experiment, in particular, why gender effects were stronger when number mismatched. This leads me to assume that this interaction supports the encoding interference hypothesis more than a purely retrieval-driven hypothesis for two reasons. First, if the retrieval mechanism were truly driving this interaction, we should have seen a main effect of Number Match, since this was the relevant feature. Secondly, the scarce literature on encoding has shown that encoding produces weaker interference effects than retrieval (Villata et al., 2018). If this is true, then perhaps we would only be able to observe encoding effects that coincide with a retrieval cue when there is no interference caused by the retrieval cue, which is what this interaction shows. This is purely speculative, and unfortunately, the lack of number match effects make any conclusions hard to reach for this interaction.

3.8.5 Comprehension Questions

Experiment 3 also contained comprehension questions after ½ of the items. In this study, there is a marginal main effect for Number Match, showing that questions following sentences with Number Matches between the target and distractor were answered less accurately ($p=0.07$). This effect is interesting because Number Match did not cause a reading time slowdown in any region, yet it affected comprehension accuracy.

	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	94% (3)	98% (2)	-4%
Gender Mismatch	92% (3)	98% (2)	-6%
Gender Match Penalty	-2%	0%	

Table 17: *% Correct on Comprehension Questions (Means and Standard Errors) - Number X Gender*

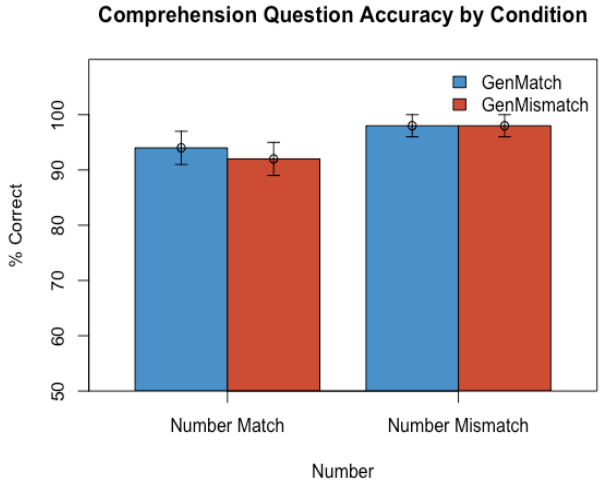


Figure 11: *% Correct on Comprehension Questions - Number X Gender*

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.968	0.780	5.089	<0.001
genderGenMatch	0.057	0.309	0.185	0.853
numberNumMatch	-0.566	0.313	-1.807	= 0.07
genderGenMatch:numberNumMatch	0.036	0.312	0.116	0.908

Table 18: *GLMER Results- Gender + Number ORC Comprehension Questions*

3.8.6 Discussion

There is no doubt that this experiment did not provide as clear of results as expected. Crucially, the predicted main effect of Number Match was not present in this the reading times and only marginally surfaced in the offline comprehension questions. Since number was the only relevant feature in these constructions, and Experiment 1 provided evidence that Number Matches do indeed disrupt processing in the form of a marginal trend, a slowdown for match conditions was strongly predicted. This effect should have also been stronger due to the grammar of Portuguese, as Number Match and Number Mismatch conditions were forced to have the same morphological variation as was observed in Experiment 1. In the mismatch case, the matrix verb and RC verb do not share the same morphological features, which completely eliminates the possibility of the RC internal subject being a competing item (see sentence below). Since the RC subject could not be a

subject of the matrix verb and has already undergone RC-internal subject-verb agreement, the mismatch conditions should have been much easier to process and made the number match penalty even more marked.

Number Mismatch - Gender Match	O homem que os meninos viram na segunda-feira dá aula de violão no centro da cidade. The man-M.SG [who the boys-M.PL. saw -PL.] on Monday teaches-SG. guitar classes in the center of the city.

Figure 12: *Sample Item - Number Mismatch - Gender Match (Number X Gender ORCs)*

However, there is another approach that can explain the lack of overt number interference in this experiment. Although it was assumed that the grammatical variation due to the verb morphology would strengthen the interference effects caused by number matches, subjects may have developed a strategy in this experiment that allowed them to ignore the baseline ORC interference. Since subjects only saw object-extracted relative clauses, they may have used shallower processing of these structures because there was no alternative and/or easier structure to consider. The existence of Trial effects does show that sentences were read faster throughout the experiment, so perhaps this structure became easy enough throughout the experiment that interference effects were no longer able to surface. Since Trial was treated as an additive factor in this experiment, no interaction between Trial and the other predictors was available in the model.

This line of thinking helps to explain the lack of number interference but does not completely explain the results, as the interaction between Gender and Number showed that that subjects did not completely ignore morphosyntactic information. As a reminder, this interaction showed that number match only caused a slowdown in cases where gender did not match. A potential interpretation of this interaction is that subjects initially pay more attention to gender than to number, but when gender cues are not in competition, participants become more attuned to number information. Essentially, if there is no overt encoding interference from gender, subjects

begin to consider the other grammatical cues. Unfortunately, this interpretation does directly align with any of the hypotheses presented in the introduction because the cue-relevancy hypothesis would have predicted a constant number effect and the cue-additivity hypothesis would have predicted an interaction when both features matched.

Before discussing the gender results and their implications, I would like to put forth a speculation about why number effects did not surface in this study and were only marginal in Experiment 1: experientially-driven cue-weighting. As proposed early on, there is likely a weighting system for cues that can explain why number match penalties surface in ORCs but not in SRCs. In its simplest formulation, this could assign higher weights to syntactic features such as Case higher than morphological agreement features due to the dialectal improbability of number features consistently surfacing (see upcoming paragraph). This weighting system only allows weak morphological interference to surface in cases where syntactic features match. With that in mind, it is also possible that this hierarchy can be extended to different classes of cues and may depend on language-specific preferences.

This study was conducted in the interior region of São Paulo, which is known for a socially stigmatized dialect of Portuguese known as *português caipira*. One of the most defining features of *caipira* dialect is a lack of number agreement, both between nouns and adjectives, as well as subjects and verbs, which is most prevalent in spoken language (Azevedo, 1984). In the case of nominal agreement, this dialect allows for the dropping of the plural marker /-s/ on the noun, as seen in the examples below. It is oftentimes the case that only the definite article is marked for plurality.

(18)

(a) NP Agreement - Standard Portuguese:	Os	carros	
	The$_{-PL.\ MASC}$	cars-$_{Pl\text{-}MASC.}$	
(b) NP Agreement - Caipira Portuguese:	Os	carro	
	The$_{--PL.\ MASC}$	cars$_{-\ SG.\ MASC.}$	
(c) Subj-Verb Agreement - Standard Portuguese:	Os	carros	vão...
	The$_{-PL.\ MASC}$	cars-$_{Pl\text{-}MASC.}$	go$_{-\ 3rd.\ Pl.}$
(d) Subj-Verb Agreement(A) - Caipira Portuguese:	Os	carros	vai
	The$_{--PL.\ MASC}$	cars$_{-\ PL.\ MASC.}$	go$_{-\ 3rd.\ Dg.}$
(e) Subj-Verb Agreement(B) - Caipira Portuguese:	Os	carro	vai
	The$_{--PL.\ MASC}$	cars$_{-\ SG.\ MASC.}$	go$_{-\ 3rd.\ Sg.}$

As the examples show, the presence of an overt number marker is not required in this dialect. Even though this is a stigmatized production, it is a well-documented sociolinguistic phenomenon (Azevedo, 1984). This information adds an interesting point to the cue hierarchy topic: Can a cue's strength change cross-linguistically? If this were the case, we could have evidence for why the RCs showed such weak effects for number matches.

If we continue under the assumption that cue-weighting is language-specific, or at least that it contains an experientially driven component, number may simply not be a relevant cue in this dialect. Although the experiment was conducted using written Portuguese, which prescriptively does not allow for the /-s/ dropping, the participants in this study are accustomed to interpreting Portuguese without relying on number agreement throughout the sentence. Even though number must be minimally marked somewhere within the DP, typically on the determiner, this dialect of Portuguese differs from more standardized dialects or even languages like English that require number to be overtly marked in all cases.

Stepping away from the number effects, the effect of gender in this experiment is relatively informative. To begin, in the RC region, there was a trend towards a main effect of gender match. These findings align with Experiment 2, which also showed that Gender features also result in a processing penalty in both SRCs and ORCs that cannot be explained by traditional retrieval

models. The consistency of these results provides strong support that gender penalties are caused something other than retrieval interference because they are not relevant for resolving the grammatical dependency in question, assuming that these are the only features that generate interference. In addition to the main effect, the pairwise comparison of the model predictors showed that gender match had the strongest effect when number was not matching. In fact, the Gender Match + Number Mismatch condition was the slowest condition in the RC region. This is an interesting finding because it truly contradicts the predictions of a cue-additivity model where both relevant and irrelevant cues contribute to interference.

The last puzzle of these results is *why* gender and number seem to interact. It is clear that both features do indeed cause a slowdown in certain conditions, but the relationship between the two seems odd. When looking at the results of all three experiments, I seem to have gathered support for the Relevant Cue Retrieval hypothesis and the Encoding hypothesis, but neither one of these would predict the way in which gender and number matches only cause penalties when the other feature is mismatched. This could potentially be explained by combining shallow processing with weak number cues, but more research is needed to directly address this question. Although it is possible, I hesitate to propose an explanation for this interaction due to the lack of number match effects in this study. If relevant cues did not cause interference, then it is likely that this interaction is due more to an experimental artifact than actual processing patterns.

3.9 Conclusion

In conclusion, the three experiments presented in this chapter generated both theoretical insights and a fair share of puzzles for future investigation. Experiments 1 and 2 demonstrated that Brazilian Portuguese speakers struggle with ORCs more than SRCs at the relative clause region. Experiment 1 also showed that Number Match makes ORCs harder to process in the post-RC

region, albeit only in the form of a marginal trend, Experiment 2 demonstrated that gender matches cause a general penalty that is likely unrelated to retrieval, and Experiment 3 supported the claims made in Experiment 2 about the role of irrelevant feature matches during sentence processing, as well as raised questions about the lack of number interference when participants only saw ORCs. What these results mean with respect to the hypotheses is rather unclear.

To begin, let's address the Cue (Super) Additivity versus the null Generalized Feature Retrieval hypothesis. In Experiment 2, there was an interaction between Clause Type and Gender Match, which suggested that the retrieval mechanism was sensitive to gender interreference. The effects led me to the following question: Since number was matched in Experiment 2, were these effects merely left-over number effects or were they generated by gender? Since the results were not replicated in Experiment 3 when both features matched, it is safe to assume that the interaction seen in both experiments was due to the number feature shared between the two. If these effects in Experiment 2 were truly additive, we should have seen this pattern emerge in Experiment 3 when both features were directly manipulated in ORCs. If the (Super) Additivity Hypotheses were correct, the Number Match-Gender Match condition should have been the hardest to process. This was not the case. Moreover, regarding the Generalized Feature Hypothesis, none of the experiments showed identical behavior for both gender and number. For example, gender matches generated main effects of clause type and number did not, and this can be taken as evidence against the Generalized Feature Retrieval Hypothesis.

With these results in mind, I would like to address the original hypotheses for this chapter beginning with the Relevant Feature Retrieval and Encoding approach. The Relevant Feature hypothesis closely aligns with the predictions made by current cue-based retrieval models, and only differs in the sense that it overtly specifies that the cue in question must be relevant to the

grammatical dependency requiring retrieval. This hypothesis is largely supported by Experiments 1 and 2. In Experiment 1, the trend for an interaction between Number Match and Clause Type showed that number feature matches were able to modulate the difficulty of ORC processing. This closely replicates the findings of Gordon et al. (2001). In addition, the same interaction was found in Experiment 2, when both Gender and Number overlapped. Since there is no evidence that the interaction in Experiment 2 was driven by Gender, it appears that this interaction represents the baseline interference levels caused by Number Match. It is odd, however, that we did not see a main effect of Number Match in Experiment 3 if this was the only mechanism at play. In Experiment 3, Number Match had a marginal effect on comprehension question accuracy, but there were no reading time effects. In fact, the interactions found in Experiment 3 were largely driven by Gender Matches being harder to process in Number Mismatch cases. Although Number Match had a numerical penalty in Gender Mismatch cases, this feature was certainly not the driving factory behind the processing delays found in this study. I will explore this concept further in the following chapter on Sluices to see if this trend is specific to RCs or if it extends to other constructions before I make any claims about why this tradeoff happened.

Finally, I return to the Encoding Hypothesis, which is likely to be the largest theoretical contribution of this chapter. Although none of these experiments can truly show the location of encoding interference in the sentence given the segmentation used, the presence of processing penalties caused by non-retrieval features is undeniable. In Experiment 2, Gender Match conditions incurred a processing penalty that was not related to clause type. As discussed earlier in this chapter, we only expect to see retrieval interference effects in the ORC constructions due to the position of the distractor and the increased number of features that it shares with the target. Unlike retrieval interference, encoding interference occurs as information is being stored, not as it

is being retrieved. Therefore, any feature that overlaps between a given item and a subsequent item, regardless of its involvement in the grammatical dependency, can cause encoding interference. As further support that the gender effects were not a spurious effect in Experiment 2, Experiment 3 tested only ORCs and still found a Gender Match penalty in the RC region. Similar findings were shown by Villata et al. (2018) in their study of Italian relative clauses, and I hope that this cross-linguistic evidence strengthens their argument in favor of an encoding interference playing an integral role in sentence processing. Although the exact mechanism that accounts for these interference effects is still under consideration (see Chapter 2), the evidence from Experiments 2 and 3 clearly provides evidence in favor of a secondary memory-based process that is largely independent from the retrieval process itself.

Although these experiments could not resolve all of the open questions about the nature of the retrieval model, by narrowing the scope of its behavior to relevant cues that are syntactically available, I have made progress towards better understanding how memory and language interact during sentence processing. The upcoming chapter further addresses these questions using a new construction, with the hope of answering some of the open questions generated by these experiments.

Chapter 4

4.1 Background

Successful sentence processing is dependent both on the storage of linguistic input and the accuracy with which it can be retrieved in order to resolve non-adjacent grammatical dependencies. In a cue-based retrieval system, all potential antecedents for a given retrieval cue are accessed through a rapid, associative cue-matching procedure. Much of the previous research on antecedent retrieval focuses on grammatical dependencies that contain overt cues, such as the subject-verb dependencies discussed in the previous chapter. In this study, we expand the discussion of cue-based retrieval beyond well-documented dependencies by examining sluiced sentences in Brazilian Portuguese.

Following Harris' (2015, 2019) studies on English sluices that showed that structural/locality information modulates similarity-based interference effects, we hope to provide evidence that the memory mechanisms involved in the processing of ellipsis are aligned with those used to resolve overt dependencies. Similar findings between two very distinct types of syntactic constructions would strengthen the claim made in Chapter 3 that retrieval and encoding are always involved in sentence processing to some extent. Moreover, both sluices and RCs contain known processing biases that are affected by the position of the target within the sentence, which can be manipulated to examine how and if these preferences modulate the degree of similarity-based interference. By demonstrating that processing preferences are able to strengthen interference effects, a deeper discussion related to how antecedent selection is constrained during retrieval can be presented. If processing preferences do not modulate the interference caused by morphosyntactic features, then it could signify that the retrieval mechanism does not consider these external biases in antecedent retrieval. The later perspective, however, is less likely to surface

because these preferences have previously been shown to affect the patterns of interference in sluiced sentences (e.g., Harris 2015, 2019).

In sum, by using Portuguese and addressing the morphological features of number and gender, a clear comparison for the behavior of these features and their role in retrieval and encoding can be made with the RC study. It is then possible to make speculations about the theoretical implications of (a) a retrieval mechanism in which all morphological cues impact the retrieval process equally and (b) a retrieval mechanism where only morphological features that are required to resolve a grammatical dependency are paired directly with a retrieval cue and (c) a system in which there are multiple sources of processing slowdowns, potentially as a result of encoding interference.

4.2 Cue-based Retrieval and Similarity-Based Interference

As discussed throughout this dissertation, the parallel cue-matching procedure in most retrieval models is sensitive to competition between items in memory. This competition is believed to be a result of the degree of overlap between the retrieval cue and the items encoded in memory. When several items in memory have overlapping features, the retrieval mechanism becomes cue-overloaded (Watkins and Watkins, 1975, 1976). For example, if there are several plural nouns in the sentence and the retrieval cue is seeking to match with a plural target, the matching procedure between the retrieval cue and its intended target will be less accurate than a sentence that only contains one plural noun and the same retrieval cue. Cue-overload then leads to similarity-based interference, which manifests itself as processing penalties in online and offline studies (see Van Dyke and McElree, 2011 for an overview of experiments).

Although the concept of cue-based retrieval has been applied to many common dependencies, such as subject-verb dependencies, few researchers have considered its application

in ellipsis structures. Much like a verb seeking a subject, ellipsis structures contain a remnant of an elided clause that must be paired with an overtly expressed correlate. In the example below of a sluiced sentence, the remnant *who* must be associated with the correlate *someone* for the sentence to be understood.

(1) Lily hugged someone, but I don't know who.

Although the overt syntactic structure of the second clause has been elided (*but I don't know who [Lily hugged]*), a handful studies on sluicing have considered how the retrieval mechanism might be involved in the processing of these sentences (Harris 2015, 2019; other ellipsis Martin and McElree, 2008). In fact, sluices appear to be sensitive to similarity-based interference in a way that mirrors the type of retrieval employed in subject-verb dependencies, albeit with distinct syntactic biases involved (section 4.1). Throughout this chapter, I will explore the nature of sluices and this correlate-remnant pairing procedure, as well as test the morphosyntactic features of gender and number in this process.

Moreover, sluices allow us to address one of the most important open questions in cue-based retrieval theory: What constrains the information can serve as a retrieval cue? In Chapter 3, I proposed a certain type of cue-weighting where overlapping syntactic information (Nominative case in the matrix and embedded subjects) was assumed to be the primary reason that ORCs experience more interference effects than SRCs. Without this Case feature overlap, no interference would be expected from morphological information, such as gender and number. This hierarchy ultimately led to the prediction that relevant features would result in an interaction with ORCs but no main effects. Sluices, on the other hand, possess a different type of structural bias that is not necessarily driven by syntactic cue overlap.

As will be discussed in section 4.3, comprehenders show a strong bias towards associating the object of a sluice with the wh-remnant. This bias is strong enough that comprehenders may be tempted to ignore grammatical information so that they do not have to associate the remnant with non-local correlate (Harris 2015, 2019). Looking at the sluices and RCs, both constructions have independently motivated processing biases. Non-local sluices are harder than local sluices not because the NPs overlap in Case features, as in RCs, but rather because a locality-driven processing bias exists for these structures. However, the source of this bias may not be *only* structural in nature, as prosodic and discourse information can modulate its strength (see Carlson et al., 2009; Harris and Carlson, 2018; Harris, 2015, 2019). With that being said, it is unclear how this type of bias fits in standard models of cue-weighting (e.g., Parker, 2019; Van Dyke and McElre, 2011). On the surface, it seems that while this bias may interact with relevant cues, we should minimally observe main effects of relevant cue matches because they are the most highly weighted morphosyntactic information in this dependency. This was not predicted for the relative clause experiments in Chapter 3.

Addressing the role of both morphosyntactic cues and processing biases that appear to derive (at least to a certain extent) from structural biases is crucial in developing a coherent theory of cues. In fact, a large debate revolves around the degree to which the grammar constrains retrieval cues. There are two main theoretical approaches to how antecedent retrieval is constrained: structure-based accounts and unconstrained-cues accounts (structure-based: Nicol and Swinney, 1989; Sturt, 2003; Xiang et al., 2009; Chow et al., 2014; constraint-based Badecker and Straub, 2002; Lewis and Vasishth, 2005; Chen et al., 2012; Jäger et al., 2015 in Harris, 2015). Simply stated, under the structure-based accounts, the grammar limits the set of potential antecedents for a retrieval cue to those in syntactically licit conditions, whereas the latter claims that every possible

type of constraint, such as focal attention or discourse status, can influence the retrieval process and allows for ungrammatical antecedents to be considered by the mechanism. Although both approaches rely on tree-geometry, in part, for limiting the potential set of antecedents, the structural account claims that retrieval interference should be greatly reduced or even entirely absent from structurally inaccessible positions, whereas an unconstrained cue account would allow interference to arise from ungrammatical antecedents if they have high enough activation from other features.

A third option, that was discussed initially in Chapter 2, is that structural constraints are weighted slightly higher than other linguistic cues, which aligns more closely with the unconstrained cue approach than the structurally-constrained retrieval perspective (e.g. Badecker and Straub, 2002; Parker et al., 2017 for discussion; Parker, 2019). One of the challenges of the debates on structural or unconstrained retrieval is that a vast majority of the research has focused on anaphor resolution, which undergo distinct syntactic and comprehension processes when compared with other types of dependencies. The benefit of adopting a *structure-first* weighting approach to retrieval constraints is that it can account for findings where only structurally accessible antecedents generate interference effects, but it also allows for the flexibility to account for interference from ungrammatical antecedents, so long as they overlap in features with the retrieval cue. Another possible benefit of this approach is that it can help to describe cross-linguistic variation. Although that structural weights will be assigned in conjunction with the dependency being processed, languages that all for some degree of grammatical variation might show different strengths for these cues and their relationship with other cues. This relates back to the discussion regarding the influence of dialect on RC processing in Chapter 3, where the optionality of number expression might make it less diagnostic or weighted lower than other

features that are more strongly imposed by the grammar. Overall, the weighted adaptation of the unconstrained cue approach can generate the broadest explanation of the way that structure interacts with other features and offers the most flexibility for adaptation to new scenarios. For example, a weighted constraint approach could easily connect with the underlying assumption that processing biases will interact with morphosyntactic features for sluices and relative clauses, assuming that the preferred and dispreferred structures will be uniquely weighted.

Although this debate is crucial in the development of cue-theory, Jäger et al.'s (2015) analysis of this topic brings up another important point: the predictions for an unconstrained cue approach are nearly identical to the predictions of encoding interference (examples cited in Jäger include Dillon (2011) and Dillon et al. 2013 as further support on how this applies to anaphors and reflexives). Unconstrained cue accounts predict that *any* form of feature overlap will produce similarity-based interference, even those that are not directly involved in retrieval. A challenge with adopting this perspective is that it does not account for the existence of encoding effects and places all interference under the umbrella of retrieval.

4.3 Sluicing and Locality

Sluicing and other forms of ellipsis have generated a large amount of interest in sentence processing studies (Dickey and Bunger, 2011; Frazier and Clifton, 1998; Harris, 2015; Harris and Carlson, 2016; McElree and Martin, 2008, 2009; Martin and McElree, 2011; Poirier et al., 2010; Yoshida et al., 2009). Sluicing is a typologically common form of ellipsis, which makes it ideal an ideal tool for comparing retrieval behavior in languages with different structural and morphological systems (Merchant, 2001). Following Merchant (2001), we assume that a sluiced sentence is formed via *wh*-movement, where the *wh*- remnant moves to the Spec of CP and the IP is elided. As a result, the surface structure contains two clauses: the second clause contains the wh-

remnant that remains after clausal ellipsis and the first clause contains the correlate, or antecedent, of the wh-remnant. In the example below, *something* serves as the correlate for the *wh*-expression *what*.

(2). a. Sam bought something at the store, but I don't know [$_{CP}$ what $_{t}$ [$_{IP}$ [~~he bought what *t*~~]].
b. Sam bought something at the store, but I don't know [$_{CP}$ what$_{t}$ he bought [~~what$_{t}$~~].

Chung et al., (1995) observe that sluices place restrictions on the type of nouns that can serve as correlates to the *wh*-remnant (i.e. a/some in English and *algum(ns)/alguma(s)* in Portuguese (also see Barker, 2013 and Barros, 2014). Typically, definites and proper names are infelicitous as correlates unless they are followed by *else (*i.e. *Sam bought the cookies at the store, but I don't know what else*) (Chung et al., 1995). This is why there is no ambiguity in sentence (2a): *Sam* would be an infelicitous correlate to *what*. Additionally, this explains why there are distinct interpretations of sentences (3a) and (3b). Sentence 3a should be interpreted as "I don't know which cookies," whereas (3b) is interpreted as "I don't know which children" due to the presence of the indefinite correlate.

(3) a. The children ate some cookies, but I don't know which ones.

b. Some children ate the cookies, but I don't know which ones

However, indefinites are not the only felicitous correlate option for wh-remnants. Weak definites can serve also as correlates, as they do not necessarily represent a unique entity (i.e. *She went to the hospital, but I don't know which one*.) (Barros, 2014; Chung et al, 1995). Additionally, disjunctions are acceptable correlates for the wh-remnant in cases of *which* remnants (i.e. *Jimmy talked to Sally or Mary, but I don't know which one*.) (AnderBois, 2010; Harris, 2019). What unites all of the possible correlates, at least for *which* sluices, is that the correlate does not represent a unique entity and allows for the comprehender to consider potential alternatives at the point of

matching the remnant and correlate (see Harris, 2019). With that being said, any correlate that does not present a viable list of alternatives (i.e. definites and quantifiers) are semantically infelicitous.

(4) *The children ate the cookies, but I don't know which ones.

(5) *The children ate all of the cookies, but I don't know which ones.

Aside from the preference to have an indefinite correlate for the *wh*-remnant, sluices also show a strong structural bias known as the Locality Bias (discussed as the Object bias in Frazier and Clifton, 1998, Carlson et al., 2009; Locality Bias, Harris 2015), which states that the preferred correlate for the *wh*-remnant is the correlate found in the most structurally local position (in the case of sluices and other forms of general clausal ellipsis). Harris (2015) found that violations of the locality bias modulated similarity-based interference effects, thus implying that some type of informational constraint is relevant for retrieval (see Harris 2015 for discussion; Harris, 2019). I am particularly interested in testing sluiced structures in Portuguese because the language-specific morphological patterns allow us to examine a wider variety of cue-matches than English. Additionally, it has been predicted that English and Portuguese sluices behave largely in the same way, but this claim has not been empirically tested. If Portuguese speakers adhere to the Locality Bias in the same way that English speakers do, we have further evidence that correlate retrieval processes operate similarly cross-linguistically.

4..4 The Current Studies

The current set of experiments examines sluiced sentences in Portuguese to investigate the interaction between morphological cues and structural information in sentence processing. In both English and Portuguese sluices, the *wh-* remnant contains a number cue (*qual/quais* - "which one(s)" in English), which must match with its intended target. However, Portuguese, unlike other Romance languages, also allows for an additional gender marker to appear on the *wh*-remnant.

The construction in Portuguese that allows for both number and gender to be relevant retrieval cues is *quais deles/delas* ("which $_{PL}$ of them $_{MAS/FEM}$"). Importantly, the gender marker is completely optional, which allows a manipulation in which gender matches between the target/distractor but remains an irrelevant feature for the correlate-remnant pairing procedure (similar to Experiment 3 in Chapter 3). Thus, Portuguese sluices provide an opportunity to test gender and number effects in retrieval, when they are both relevant features, as well as set the stage for a discussion about encoding interference when gender is no longer a relevant retrieval feature.

The overall design of this study is largely inspired by Harris (2015), an eye-tracking study that examined sluiced sentences in English. The results of Harris (2015) provided initial evidence that the position of the intended correlate influences similarity-based interference, such that when the Locality Bias is violated (here, a subject-correlate, as opposed to an object correlate), similarity-based interference effects increase. We hope to replicate Harris' findings regarding Locality Bias violations in a non-English language, as well as examine the effects of gender and number cue matches on retrieval.

Given the findings in Harris (2015) that show that number cue matches can cause retroactive interference effects when the Locality Bias is violated, we would expect example (6) to have no similarity-based interference effects. This is because the number cue is not shared between the distractor and the probe, even though the Locality Bias is violated. In contrast, example (7) should show interference effects because the [+plural] cue overlaps between the target, the distractor and the probe and the Locality Bias is violated.

(6) **Algum menino** adotou <u>os cachorros</u>, mas eu não sei quais.
Some boy-$_{MAS.\ SG.}$ adopted the dogs-$_{MAS.\ PL.}$ but I don't know which-$_{PL.}$

(7) **Alguns meninos** adotaram <u>os cachorros</u> , mas eu não sei quais.
Some boys-$_{MAS.\ PL.}$ adopted. the dogs-$_{MAS.\ .PL.}$ but I don't know which-$_{PL.}$

What differs between Portuguese and English is that the target and distractor can also match in gender. For example, in Portuguese, the wh-remnant can optionally be *quais deles/delas,* translating roughly to "which ones of them-$_{MAS/FEM}$", where "them" is marked for gender. Since the *wh*-remnant can overtly contain a gender marker, most models of cue-based retrieval would predict that gender will behave very similarly to number at the point of retrieval, as it is employed in the dependency being resolved. Examples like (6-7) allow us to primarily evaluate the effects of gender on retrieval because the number must always be plural, which makes gender the only cue that can be experimentally manipulated. In both examples (8-9), the Locality Bias has been violated; however, we only expect to see retroactive interference effects provoked by gender cue matches in example (8), where *deles* overlaps with both the masculine target and distractor. Using the structures presented in (6-7) and (8-9) will allow us to compare the roles of these unique morphological cues at the point of retrieval, which has yet to be done in the literature on this topic.

(8) **Alguns meninos** adotaram <u>os cachorros</u> , mas não sei quais deles.
Some-$_{M.PL}$. boys$_{M.PL}$ adopted the dogs$_{M.PL.}$, but I don't know which ones of them$_{M,PL.}$

(9) **Algumas meninas** adotaram os cachorros , mas não sei quais delas.
Some-$_{F.PL}$. girls$_{F.PL.}$ adopted the dogs$_{M.PL.}$, but I don't know which ones of them$_{F,PL.}$

4.4.1 Hypotheses and Predictions

The primary hypotheses in this chapter are identical to those presented in Chapters 2 and 3, so I will not repeat their definitions here. As discussed in Chapters 2 and 3, there is an underlying assumption is that processing preferences will modulate the strength of interference effects, resulting in an interaction between the two factors (similar to ORC/SRC distinction and supported by Harris, 2015, 2019). Although this leads to the expectation of an interaction in the upcoming predictions, it is also possible that only two main effects will be observed: a penalty for non-local targets and a penalty for feature matches between the target and distractor. Two main effects, as

opposed to an interaction, would indicate that the processing preference for local sluices is not strong enough, or perhaps not weighted higher, than morphosyntactic feature matches in Portuguese. This second option, although a possibility, is less likely given the preexisting sluicing and RC evidence that dispreferred structures increase similarity-based interference effects related to other linguistic content (e.g. RC – Gordon et al., 2001; sluices – Harris, 2015, 2019).

Relevant Feature Retrieval,* predictions*: Experiments 1 and 2 deal directly with relevant retrieval features. In Experiment 1, number is the only relevant feature, and in Experiment 2, gender is the relevant feature. Therefore, in both experiments, the Feature Match – NonLocal condition should be the hardest to process, as there is a higher likelihood that retroactive interference will surface in this condition (Harris 2015, 2019). Similar to the predictions presented for Number Matches in Chapter 3, I do not necessarily expect to see a main effect of number or gender match in Experiments 1 and 2. This is because the existing processing literature on sluicing has shown that interference effects primarily arise in the Non-local sluices only.

If a general penalty for Feature Match were to be observed, the effect would still be insightful. This could be a result of proactive interference effects surfacing in the local sluices, which tend to be weaker in processing studies (McElree et al., 2003) or perhaps even residual effects of encoding interference. Both of these will be discussed in greater detail, should a main effect of feature match appear in Experiments 1 or 2.

***Generalized Feature Retrieval,* predictions:** The Generalized Feature Retrieval hypothesis could be supported by Experiment 3. Given that Experiments 1 and 2 only manipulate relevant features, it is impossible to identify how the mechanism reacts to irrelevant features. In Experiment 3, Gender is no longer relevant to the correlate-remnant pairing procedure. Therefore, if we see that gender still interacts with Locality and produces the same type of interference effects that number

does, we would have support that the retrieval mechanism is sensitive to all overlapping cues. This hypothesis was largely unsupported in Chapter 3 and it is not expected to be supported in this chapter. It is presented merely as a logical alternative for the Relevant Feature Retrieval hypothesis.

***Cue (Super)Additivity,* predictions:** The Cue (Super)Additivity hypothesis predicts that retrieval interference effects will be stronger when the target and distractor overlap on more than one cue. If this were the case, Experiments 2 and 3 are where these effects would emerge. In Experiment 2, although the focus is on the relevant feature of gender, number features also overlap between the target and the distractor due to grammatical constraints of Portuguese sluices. Therefore, if the effects in Experiment 2 are much stronger than those in Experiment 1 (where only number feature overlaps between the target and distractor), this could be interpreted as evidence for cue additivity. It is important to note that this experiment differs from the gender experiment in Chapter 3, because in that experiment, gender was not a relevant cue. It is unclear how the retrieval mechanism will process *two* relevant cues that both overlap. Finally, Experiment 3 more directly addresses this question by crossing Number and Gender directly. In Experiment 3, if the Gender Match – Number Match condition is significantly harder to process than the other variations of matches between these two features, this could be construed as evidence for cue additivity.

***Encoding*, predictions:** One of the primary goals of these experiments was to search for potential encoding interference effects in Portuguese sluices. Experiment 3 was designed to specifically test this hypothesis. In Experiment 3, number is the only relevant feature for retrieval, but gender features are also manipulated between the target and distractor. As in Chapter 3, any effect of gender, particularly in the form of a main effect that is not sensitive to clause type or sentence

structure, can be interpreted as interference effects when the feature is not required to resolve the dependency.

4.4.2 General Method and procedure

The three experiments reported in this chapter were conducted using self-paced reading and were used as fillers for the Experiments in Chapter 3. In addition to reading, participants were also asked to identify the correlate of the wh-remnant in a comprehension question immediately following the sentence. All items were presented using the program Linger (Rhode, 2002) in a quiet room using Macintosh computers.[11] The sentences were segmented identically across experiments for ease of comparison. A sample item with segmentation (marked by "/") and comprehension questions is presented below.

Sentence:	Na feirinha de adoção/as vendedoras/viram/algumas veterinárias/mas eu não posso dizer/ quais delas/ porque eu tive que sair cedo/ do evento.
	At the adoption fair, /the saleswomen-$_{\text{FEM. PL.}}$/ saw some veterinarians-$_{\text{FEM. PL.}}$, / but I can't say which ones of them-$_{\text{FEM. PL.}}$/ because I had to leave/ the event early
Question:	Qual é a informação que eu não posso dizer? What is the information that I can't say?
Answer:	1. Quais vededoras 2. Quais veterinárias 1. *Which saleswomen* 2. *Which veterinarians*

Figure 13: *Sample Item, Segmentation and Method for Sluices*

University of Campinas and the Pontifical University of Campinas.[12] For Experiments 1 and 3, the participants were compensated via participation in a raffle where they were given one

[11] Experiment 2 experiment differs slightly in this procedure, as participants were also asked to provide an acceptability judgement about the entence.

[12] Only applicable for the Gender Experiment. The Gender Experiment was conducted in 2019 and the remaining two experiments were conducted in 2019.

ticket for their participation and an additional ticket for each participant that they invited for the Number and Number + Gender Experiments. In Experiment 2 (conducted during a separate fieldwork session), participants volunteered without any formal compensation.

4.4.3 Design

Experiments 1 and 2 (Number and Gender) both employed a 2X2 design in which Feature (Match, Mismatch) was crossed with Locality (Local Target, NonLocal Target). In both experiments, the target was always an indefinite NP and the distractor was an infelicitous correlate for the wh-remnant (definite NP in Number experiment, definite or quantified NP in Gender experiment). The Gender X Number Experiment differed from the other two by using a between-subjects 2X2 design. One set of subjects saw local sluices with a Gender (Match, Mismatch) X Number (Match, Mismatch) manipulation and the other set saw the same Number X Gender design for Non-local sluices. The results of both sub-experiments were combined in the final analysis of the experiment to determine how the three factors interacted.

4.4.4 Materials

In all 3 experiments, subjects were presented with 24 experimental sluices and an appropriate number of non-experimental items (RCs) and filler sentences to maintain a 1:4 ratio between critical and non-critical items. The nature of the filler items differed slightly in each experiment, and these variations will be discussed in subsequent sections. Importantly, all of the sluicing items contained identical regioning to maintain consistency across the experiment and items.

The selection of nouns used in this experiment is identical to the process discussed in Chapter 3. As a brief reminder, all of the nouns used in this experiment were animate profession nouns, due to their gender flexibility in Portuguese. The majority of nouns in this class can shift gender by simply alternating the gender morpheme (e.g. *médic-o* "doctor$_{M.}$" versus *médic-a*

"doctor$_{F.}$"). As part of the lexical characteristic controls, I opted to use the average length and frequency of both the masculine and feminine forms of each noun in paired t-tests testing the similarity between these factors. All targets and distractors were then matched on length (target M = 8.67, SE =0.46; distractor M = 8.04, SE=0.47) and frequency (target M =410.13, SE=122.84 ; distractor M = 516.85, SE=160.05) prior to testing in paired t-tests, p's > 0.2.

As in the RC study, half of the items were presented in their preferred gender and half were presented in their dispreferred gender. This means, a word like *enfermeira* ("nurse-$_{F.}$"), which has a feminine bias, was presented half of the items, and it was presented in masculine form in the other half of the items: *enfermeiro* ("nurse-$_{M.}$").

4.4.5 General Analysis

The reading times of the final three regions of the sentence and the comprehension question results were analyzed using linear and logistic mixed effects regression models, respectively. The analysis was conducted using the lme4 package (Bates and Maechler, 2009) in R version 3.5.1 (R Development Core Team, 2018). As in Chapter 3, three models were considered for all of the analyses conducted in this chapter.

The hypothetically simplest condition, *Local - Feature Mismatch*, was set as the baseline for the statistical analysis using sum coding. The Baseline model contained the primary predictors and random by-subject and by-item intercepts: Feature Match * Locality for Experiments 1 and 2 and Gender * Number * Locality. Two other models containing Trial were also generated: (1) Trial as an interactive predictor (Baseline * Trial) and (2) Trial as an additive factor (Baseline + Trial). Three pairwise chi-squared tests were conducted for each experiment using the anova function in R. The first two comparisons were between the Baseline and each Trial model and a final comparison was conducted between the two trial models. There were slight variations between

experiments regarding model fit, so the model of best fit is reported individually for each experiment.

In addition, outliers were removed from the analysis using winsorization, a method where scores above the 95th percentile or below the 5th percentile are transformed to the score at the 95 percentile and 5th percentile in order to avoid data loss (Dixon, 1960; Tukey, 1962).

4.5 Experiment 1: Number cues

Experiment 1 was conducted to examine the role of number features when relevant for correlate-remnant pairing procedures. This first experiment is considered the baseline for all of the upcoming experiments (Gender, Gender X Number) and also acts as a point of comparison to Chapter 3, where number was the only feature relevant to retrieval. In this study, participants were presented with 24 sluices in which the wh-remnant containing the wh-expression *which* with singular or plural marking (*qual*/*quais* – "which$_{SG./PL}$"). In this construction, number is the only relevant feature in the remnant-correlate pairing procedure.

4.5.1 Participants

A total of 33 native speakers of Brazilian Portuguese participated in this study, but one subject was removed due to a diagnosed attention disorder that they self-reported to have interfered with their participation in the study. All participants were university students at the State University of Campinas in São Paulo, Brazil. As an incentive for participation, participants received one ticket to a raffle for their participation and also received an additional ticket for each additional subject that they referred to participate in the experiment.

4.5.2 Design and Materials

Subjects were presented with a total of 24 experimental items, 32 distinct experimental fillers (20 RCs from Chapter 3), 65 nonexperimental fillers, 5 catch items, and 4 ambiguous sluices to

maintain a roughly 1:4 ratio between sluices and non-sluices. This experiment also used a Latin square design that crossed Number (Match vs. Mismatch) by Locality (Local Target vs. NonLocal Target). The target item consistently matched in number with the wh-remnant and the number of the distractor item was manipulated. In addition, there were two non-orthogonal contrasts. First, half of the items contained a plural target and half contained a singular target. Second, the target and the distractor never matched in gender in order to prevent a potential confound from gender cues. A sample item is shown below.

Na empresa nova... At the new company...					
Object Correlate-Number Match	os	empreendedores	demitiram	algumas	funcionárias
	the $_{M,PL.}$	entrepreneurs $_{M.PL.}$	fired $_{PL.}$	some $_{F,PL.}$	employees $_{F,PL.}$
Object Correlate-Number Mismatch	o	empreendedor	demitiu	algumas	funcionárias
	the $_{M,SG.}$	entrepreneur $_{M.SG.}$	fired $_{SG.}$	some $_{F,PL.}$	employees $_{F,PL.}$
Subject Correlate-Number Match	algumas	empreendedoras	demitiram	os	funcionários
	some $_{F,PL.}$	entrepreneurs $_{F,PL.}$	fired $_{F,PL.}$	the $_{M,PL.}$	employees $_{M,PL.}$
Subject Correlate-Number Mismatch	algumas	empreendedoras	demitiram	o	funcionário
	some $_{F,PL.}$	entrepreneurs $_{F,PL.}$	fired $_{F,PL.}$	the $_{M.SG.}$	employee $_{MSGL.}$
....mas eu não posso dizer quais... but I can't say which $_{PL.}$					

Table 19: *Sample Item - Number Sluices*

4.5.3 Analysis

Three models were considered in the analysis: (1) Baseline (Number * Locality), (2) Trial as an interactive predictor (Number*Locality*Trial), and (3) Trial as an additive predictor (Number*Locality + Trial). In all regions, chi-squared tests showed that the Trial models were a better fit for the data than the baseline (p's<0.05). When comparing the Trial models, the simpler model with Trial as an additive predictor was a significantly better fit for all of the data in all regions (p<0.05). There was a marginal trend in the critical region showing that the model treating Trial as an interactive predictor was better, but the model will not be reported ($X^2(11)=6.36$, p=0.096). All of the results in this section are from the model containing Trial as an additive factor.

4.5.4 Predictions

Following the hypotheses presented in section 4.4.1, two primary effects were predicted. First, there should be a main effect of Locality, such that non-local targets are penalized in either reading times or comprehension questions. These findings would align with the studies on English discussed in the introduction of the chapter, showing that non-local sluices are subject to retroactive retrieval interference and that Portuguese contains a Locality Bias (e.g. Harris, 2015). Second, an interaction showing that the Number Match – NonLocal condition is the hardest to process should arise. This interaction not only aligns with the Relevant Feature Retrieval hypothesis that is discussed throughout this dissertation, but also with the overarching assumption of this dissertation that antecedent retrieval can be constrained by structural preferences. This interaction would show that positional information (the position of the target) influences the retrievability of a target when other relevant features (e.g., number) also overlap. As with the number experiment in Chapter 3, Number Match should not necessarily be penalized in both local and non-local sluices (a main effect of match), as previous literature has shown that these effects only surface when the Locality Bias is violated (Harris 2015). The absence of an interaction would simply imply that the Locality Bias does not increase the interference caused by morphosyntactic features, but two main effects would demonstrate that 1) Portuguese contains a Locality Bias and 2) number matches generate interference effects.

4.5.5 Results

The final three regions of the sentence and the comprehension questions were analyzed using linear mixed effects regression models using the lmer4 package in R version 3.5.1. The results for the LMER analysis can be found in Table 20.

	Critical Region		
	Local	**NonLocal**	**NonLocal Penalty**
Match	887 (51)	1007 (65)	120
Mismatch	728 (30)	849 (38)	121
Match Penalty	159	158	
	Spillover Region		
	Local	**NonLocal**	**NonLocal Penalty**
Match	870 (36)	980 (42)	110
Mismatch	774 (25)	1005 (49)	231
Match Penalty	96	-25	
	Final Region		
	Local	**NonLocal**	**NonLocal Penalty**
Match	1047 (64)	1063 (44)	16
Mismatch	779 (31)	830 (37)	51
Match Penalty	268	233	

Table 20: *Means and Standard Errors for Number Experiment*

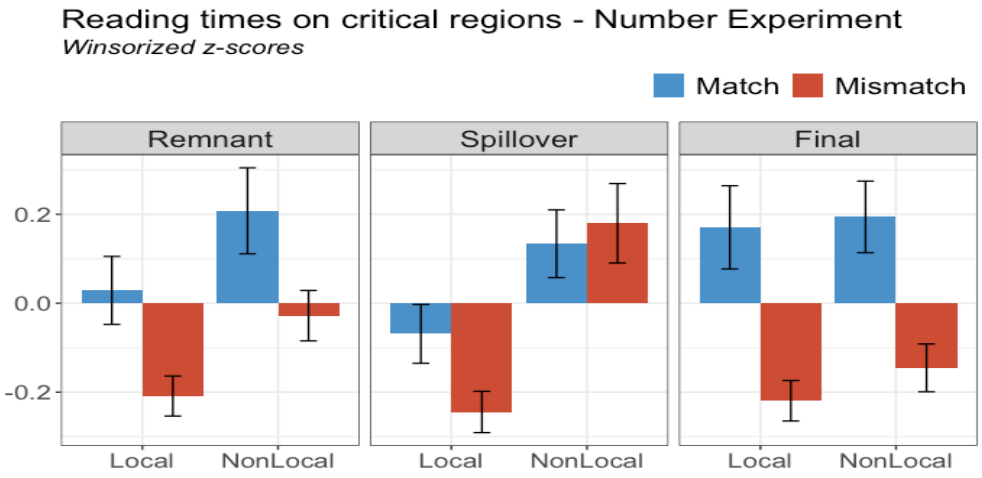


Figure 14*: Standardized Reading Times for Number Sluices*

Critical Region	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	942.536	77.938	34.561	12.093	<0.001
NonLocal	43.607	13.478	708.706	3.235	<0.01
Number Match	48.913	13.478	708.678	3.629	<0.001
Trial	-11.96	5.371	36.203	-2.227	<0.05
NonLocal:Number Match	-8.059	13.48	709.222	-0.598	0.550
Spillover Region					
Intercept)	961.406	49.684	37.570	19.350	<0.001
NonLocal	71.484	12.643	707.666	5.654	<0.001
Number Match	10.582	12.643	707.653	0.837	0.403
Trial	-8.492	3.209	30.357	-2.646	<0.05
NonLocal:Number Match	-21.184	12.644	707.898	-1.675	= 0.094
Final Region					
(Intercept)	1054.392	78.875	33.473	13.368	<0.001
NonLocal	19.155	15.505	704.657	1.235	0.217
Number Match	104.588	15.505	704.633	6.746	<0.001
Trial	-15.038	5.231	29.624	-2.875	<0.01
NonLocal:Number Match	-2.211	15.507	705.088	-0.143	0.887

Table 21: *LMER Results - Number Sluices*

4.5.5.1 Locality

If the Locality Bias can modulate morphosyntactic interference, which is supported by Harris' studies on English sluices (2015, 2019), this would predict that non-local targets will be more challenging to retrieve due to their position in the sentence (also supported in Carlson et al., 2009; Frazier and Clifton, 1998). The present study corroborates these findings, as shown by a penalty for non-local targets in the critical region (214 ms, $p<0.01$) and the spillover region (341 ms, $p<0.001$).

4.5.5.2 Number Match

When designing this experiment, there was no overarching prediction that Number Match would cause processing penalties for both local and non-local sluices. Although number was the only relevant retrieval feature, it was expected that the strength of the retroactive interference effects in non-local sluices would override the proactive interference effects of number match in the Local

conditions. With that being said, a match penalty was found in both the critical (337 ms, $p<0.001$) and final regions (501 ms, $p<0.001$). Overall, this lends support for the Relevant Feature Hypothesis, since number was the only feature being manipulated, but it is unclear how these findings related to the Structurally Driven Retrieval Hypothesis.

4.5.5.3 Trial

Trial effects were present in every region ($p<0.05$). These effects show that subjects began to read faster as the experiment continued.

4.5.5.4 Interaction between Number Match and Locality

An interaction between Number Match and Locality would support the Relevant Feature and Structurally Driven Retrieval Hypotheses, as well as replicate the Harris (2015) study on English sluices. Looking at the winsorized means, it appears that an interaction should surface in the critical region, as there is a 241 ms penalty for non-local targets and a 317 ms penalty for Number Match. Oddly, this interaction did not become significant in any of the models. The plot below represents the numerical trends towards an interaction between Locality and Number Match, but since this effect did not reach significance, it will not be considered.

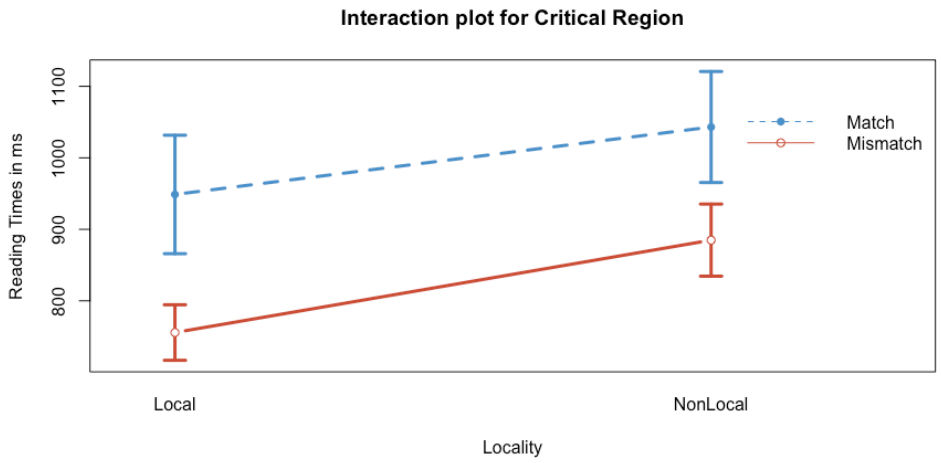


***Figure 15**: Interaction Plot for reading times in Critical Region*

4.5.6 Comprehension Questions

After each sentence, participants were asked to respond to a forced-choice comprehension question, in which they identified the correlate for the wh-remnant. The correct answer was consistently an indefinite target that matched in number with the remnant.

	Local	NonLocal	NonLocal Penalty
Match	85 (4)	42 (4)	-41%
Mismatch	98 (1)	86 (2)	-12%
Mismatch Penalty	-13%	-44%	

***Table 22:** % Correct on Comprehension Questions (Means and Standard Errors) - Number Sluices*

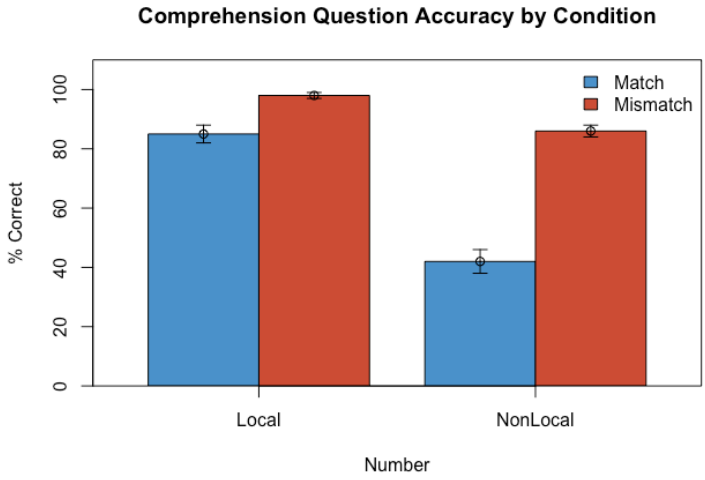


***Figure 16**: % Correct on Comprehension Questions - Number Sluices*

	Estimate	df	t value	Pr(>|t|)
(Intercept)	1.128	0.205	5.514	<0.001
Nonlocal	-0.034	0.093	-0.369	0.712
Number Match	-0.884	0.097	-9.113	<0.001
Nonlocal:Number Match	0.035	0.093	0.376	0.707

Table 23: *LMER Results, Number Sluices - Comprehension Questions*

There are main effects of Locality ($p<0.001$) and Number Match ($p<0.001$), for which the means and standard errors are presented in Table 22. There is no interaction present, which is congruent with the online study. These findings suggest that Brazilian Portuguese speakers do experience similarity-based interference cause by Number Match and are sensitive to a Locality Bias. In fact, this sensitivity is strong enough that they were actually willing to entertain and select semantically infelicitous correlates when their Locality Bias and Number Mismatch preferences were maintained.

4.5.7 Discussion

The findings of Experiment 1 suggest that Brazilian Portuguese speakers are sensitive to the Locality Bias and that number is a strong enough feature to generate similarity-based interference effects. These effects are consistent in both online and offline measures. Unfortunately, the interaction needed to truly replicate Harris' (2015) findings on English sluices was not significant in any region. At this point, there is not enough evidence to claim why these effects did not surface, since only one experiment has been conducted and we looked at slightly different remnant types in this study. A single experiment is not sufficient to speculate that Brazilian Portuguese speakers treat number features differently in processing than English speakers. It could be the case that the interaction would surface using a more sensitive method, such as eye-tracking. This discussion will be revisited throughout the chapter, as the results of related experiments are presented. For now, these findings provide early support for both the Structurally Driven and Relevant Feature retrieval hypotheses.

4.6 Experiment 2: Gender as a Relevant Feature

This experiment was conducted to verify that gender features can also generate similarity-based interference effects when they are relevant to the grammatical dependency. The Relative Clause Experiments in Chapter 3 did not allow us to test this feature, which is one of the primary motivations behind looking at sluices in Portuguese. As previously mentioned, Brazilian Portuguese allows for an optional gender marking on plural wh-remnants *quais deles/delas*, which loosely translates to "which$_{PL.}$ of them$_{M./F.}$." Since number is always matched in this construction, gender becomes a relevant retrieval feature, even though number is still expressed by the pronoun.

The results reported below are a subset of a larger experiment on gender and correlate selection conducted in Brazilian Portuguese (Lawn and Harris, in progress). In the main experiment, we analyzed reading times and comprehension question accuracy for 24 sluices, where the gender between the target and distractor were manipulated. However, this study also focused on the effect of article type on correlate selection. Therefore, 12 of the sluices contained a definite article to mark the distractor (i.e. *os medicos* – the doctors $_{PL. MASC.}$) and 12 contained a quantifier (i.e. *todos os medicos* – all $_{PL. MASC.}$ of the $_{PL. MASC.}$ doctors $_{PL. MASC.}$).[13] Since the behavior between article type differed in the larger study, I will only be reporting results for the 12 sluices that contained definite distractors, as this was the distractor type used in Experiments 1 and 3.

4.6.1 Participants

A total of 33 native speakers of Brazilian Portuguese participated in this study, but only 32 were analyzed as one participant was removed after telling the experimenter that they selected answers at random. Of the 32 participants included in the analysis, 29 were undergraduate students and 3 were graduate students from a wide variety of majors. All subjects received correctly answered

[13] This manipulation was only possible with the gender experiment, as the quantifier sub-manipulation can only be used in situations where both the target/distractor are plural.

80% or above of the filler comprehension questions and reported normal reading and attention skills.

4.6.2 Design and Materials

This experiment employed a 2X2 design crossing Gender (Match vs. Mismatch) with Locality (Local target X NonLocal target). In the main experiment, subjects were presented with a total of 24 experimental items containing a planned non-orthogonal contrast with definiteness, 52 unrelated experimental fillers (including the RC items from the previous chapter), 4 catch items, and 4 ambiguous sluices. In this section, 12 of the 24 sluices are reported, as they contained definite distractors (as opposed to quantifiers) and align with the distractor type used in the other experiments in this chapter. The target item consistent matched in gender with the wh-remnant and the gender of the distractor was manipulated to test for interference effects. A sample item is shown below.

Na empresa nova... At the new company...					
Object Correlate-Gender Match	as	empreendedoras	demitiram	algumas	funcionárias
	the $_{F,PL.}$	entrepreneurs $_{F,PL.}$	fired $_{PL.}$	some $_{F,PL.}$	employees $_{F,PL.}$
Object Correlate-Gender Mismatch	os	empreendedores	demitiram	algumas	funcionárias
	the $_{M,PL.}$	entrepreneurs $_{M,PL.}$	fired $_{PL.}$	some $_{F,PL.}$	employees $_{F,PL.}$
Subject Correlate-Gender Match	algumas	empreendedoras	demitiram	as	funcionárias
	some $_{F,PL.}$	entrepreneurs $_{F,PL.}$	fired $_{F,PL.}$	the $_{F,PL.}$	employees $_{F,PL.}$
Subject Correlate-Gender Mismatch	algumas	empreendedoras	demitiram	os	funcionários
	some $_{F,PL.}$	entrepreneurs $_{F,PL.}$	fired $_{F,PL.}$	the $_{M,PL.}$	employees $_{M,PL.}$
....mas eu não posso dizer quais delas... but I can't say which ones of them $_{F. PL.}$					

***Table 24**: Sample Item for Gender Sluices*

4.6.3 Procedure

This experiment was conducted one year before the others reported in this chapter and had a slightly different procedure. Participants still participated in a self-paced reading task using Linger (Rohde, 2002). However, the experiment was conducted on a MacBook Air using an external Rosewell RK-9100XBBR gaming keyboard to minimize the delay of button presses and provide the most accurate reaction times possible. The experiments were conducted in quiet rooms within the university (library or laboratories) with white noise (a ceiling fan and/or air conditioning unit) to block out external noise. In addition to the reading task and comprehension questions discussed in the general analysis, participants were also asked to provide a naturalness rating using a scale of 1-7. A sample procedure of this is shown below.

Sentence:	Na feirinha de adoção/as vendedoras/viram/algumas veterinárias/mas eu não posso dizer/ quais delas/ porque eu tive que sair cedo/ do evento. *At the adoption fair, the saleswomen-$_{FEM.\ PL.}$ saw some. veterinarians-$_{FEM.\ PL.}$, but I can't say which ones of them-$_{FEM.\ PL.}$ because I had to leave the event early* .
Naturalness Rating:	Avalie a naturalidade da frase anterior *Evaluate the naturalness of the previous sentence* *(completely natural)) 1 2 3 4 5 6 7 (completely unnatural)*
Comprehension Question:	Qual é a informação que eu não posso dizer? *What is the information that I can't say?*
Answer:	1. Quais vededoras 2. Quais veterinárias 1. *Which saleswomen* 2. *Which veterinarians*

Figure 17: *Sample procedure for gender sluices*

4.6.4 Analysis

The same method of analysis was used in this study as in Experiment 1. 3 LMER models were compared using pairwise anova tests for each region. In all of these comparisons, there was no significant difference in model fit between the baseline model and the models containing Trial. Therefore, the simplest (baseline) model was selected: Locality*Gender + (1 | Subject) + (1 | Subject).

4.6.5 Predictions

The primary predictions for this study were identical to those presented for Experiment 1: a penalty for violating the Locality Bias, as well as an interaction between Locality and Gender Match. If the Relevant Feature Hypotheses is correct, gender and number should present nearly identical patterns. The only slight difference between this study and Experiment 1 is that it opens the door for speculations about the Cue Additivity Hypothesis since there is a baseline of number interference across all conditions. Since both number and gender are relevant cues in this construction, even though all sentences are marked for plural, strong effects for gender in Experiment 2 than number in Experiment 1 might be indicative of additivity. Finally, I would like to point out that this study cannot provide any insight about the Generalized Feature Retrieval Hypothesis, since both gender and number are relevant, nor about the Encoding hypothesis, as no non-retrieval features are being manipulated.

4.6.6 Results

The results for all regions are presented below. Table 25 contains the results of the LMER models.

	Critical Region		
	Local	**NonLocal**	**NonLocal Penalty**
Match	1024 (60)	1071 (69)	47
Mismatch	938 (39)	960 (41)	22
Match Penalty	86	111	
	Spillover Region		
	Local	**NonLocal**	**NonLocal Penalty**
Match	1153 (45)	1220 (52)	67
Mismatch	1058 (39)	1094 (48)	36
Match Penalty	95	126	
	Final Region		
	Local	**NonLocal**	**NonLocal Penalty**
Match	1205 (65)	1374 (84)	169
Mismatch	961 (48)	878 (40)	-83
Match Penalty	244	496	

Table 25: *Reading Times for Gender Sluices in ms - means and standard errors*

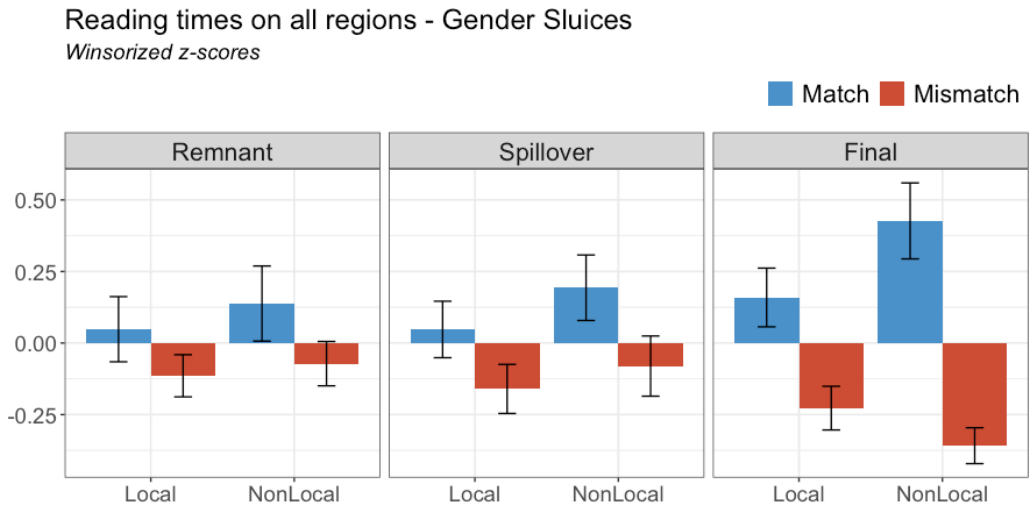


Figure 18*: Normalized Z-scores for Gender Sluices across regions*

Critical Region	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	998.221	46.867	24.351	21.299	<0.001	
Nonlocal	17.351	24.692	338.000	0.703		0.483
GenderMatch	48.986	24.692	338.000	1.984		0.050
Nonlocal:GenderMatch	6.224	24.692	338.000	0.252		0.801
Spillover Region						
(Intercept)	1131.311	54.523	23.531	20.749	<0.001	
Nonlocal	25.783	20.065	338.000	1.285		0.200
GenderMatch	55.043	20.065	338.000	2.743	<0.01	
Nonlocal:GenderMatch	7.566	20.065	338.000	0.377		0.706
Final Region						
(Intercept)	1104.710	61.630	17.070	17.925	<0.001	
Nonlocal	21.500	28.410	338.000	0.757		0.450
GenderMatch	184.850	28.410	338.000	6.507	<0.001	
Nonlocal:GenderMatch	62.910	28.410	338.000	2.214	<0.05	

***Table 26**: LMER Results - Gender Sluices*

4.6.6.1 Locality

As in Experiment 1, we expected a processing penalty when the Locality Bias was violated. Unexpectedly, there was no penalty for non-local targets in any region. The lack of effect is not particularly concerning, as only 12 sluices were analyzed, which makes it difficult to compare the effects in Experiments 1 and 2. Moreover, there is evidence that the Locality Bias influenced processing in this experiment, in the form of an interaction between Locality and Gender Match (discussed in section 4.6.6.3).

4.6.6.2 Gender Match

This experiment aimed to verify that gender features act as retrieval cues when they are relevant to the dependency. As a reminder, it was impossible to completely isolate number and gender features in this study, as the gender marker on the wh-remnant is only possible in plural constructions. This means that there may have been a baseline of interference for number matches,

but since this was consistent across conditions, any effects of gender match can be attributed to this feature alone.

There was a penalty for Gender Match in all regions (critical region, 197 ms, $p<0.05$; spillover region, 221 ms, $p<0.01$; final region, 750 ms, $p<0.001$). These effects are very consistent with the findings for Number Match in Experiment 1, which showed a penalty that also increased in strength between the critical and final regions. The similarity between these results strongly suggests that both Gender and Number have similar effects on sentence processing when they are both relevant to the dependency. These similarities between morphosyntactic features are crucial in developing a coherent theory of what types of linguistic information serve as cues and how the retrieval mechanism responds to morphosyntactic information.

4.6.6.3 Interaction between Locality and Gender Match

The Structurally Driven and Relevant Feature Retrieval Hypotheses generated the expectation that the Gender Match – NonLocal condition should be the hardest to process. The interaction appeared in the final region ($p<0.05$), lending further support to the current literature on the topic and demonstrating that Gender is just as strong as number as a retrieval feature. In fact, the interaction could be considered evidence that gender features were slightly stronger than number since the Locality – Number Match interaction did not surface in Experiment 1. However, it is more likely that this effect is a result the sentence containing two *relevant* cues than a difference between the behavior of two relevant features. Since there was a baseline of number interference, the interaction between Gender and Locality likely surfaced due to the increased difficulty of having to process two overlapping features. Although this is indeed a super-additive effect, this finding does not support the Cue (Super) Additivity Hypothesis directly, as that hypothesis is concerned with the behavior of a relevant and irrelevant cues. Initially, there were no hypotheses made about how two

relevant cues would behave when they overlapped, but these findings would warrant an extension of the Cue (Super) Additivity Hypothesis.

4.6.7 Comprehension Questions

The offline and online results are largely comparable but have one important difference. While both the methods contain a main effect for gender match (comprehension questions, $p<0.001$), the comprehension questions also show a penalty for non-local targets that was not present in the online measures ($p<0.01$). This finding is aligned with the previous literature and other experiments in this chapter show strong effects of the Locality Bias. As in Experiment 1, there is no interaction between Gender Match and Locality, but the main effects do strongly imply that similarity-based interference and target position strongly influence that interpretation that a reader assigns to a sentence.

	Local	NonLocal	NonLocal Penalty
Gender Match	61% (5)	44% (5)	-17%
Gender Mismatch	90% (3)	78% (4)	-12%
Gender Match Penalty	-29%	-34%	

Table 27: *Comprehension Question Result Accuracy - Gender Sluices*

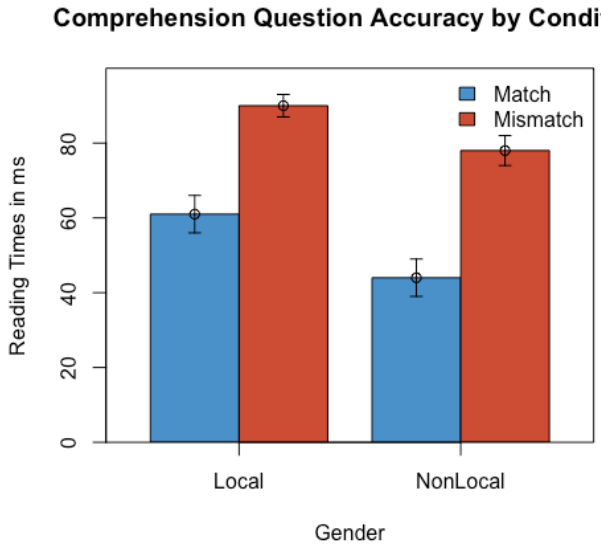


Figure 19: *Comprehension Question Accuracy Means Gender Sluices (%)*

	Estimate	se	t value	Pr(>|t|)
(Intercept)	1.128	0.206	5.514	<0.001
GenderMatch	-0.034	0.093	-0.369	0.712
NumberMatch	-0.884	0.097	-9.113	<0.001
GenderMatch:NumberMatch	0.035	0.093	0.376	0.707

Table 28: *LMER Results, Gender Sluices - Comprehension Questions*

4.6.8 Discussion

The primary findings of this experiment were (1) offline and online penalties for Gender Match, (2) an offline penalty for violating the Locality Bias, and (3) an online interaction in the final region between Gender Match and Locality. First are foremost, these findings are important because they show that, at least in BP, both gender and number can be retrieval cues in certain constructions. More importantly, now that gender effects have been observed when the feature is *relevant,* we are now able to revisit the meaning of gender match effects when the feature is irrelevant to retrieval. Any differences in behavior could foster support for the existence of two mechanisms: encoding and retrieval.

One of these questions that this study raises is: Why did the interaction between Feature Match and Locality only appear for gender features and not for number? To begin to address this question, it must be noted that a direct comparison cannot be made between the two experiments since different procedures and number of items were presented in each experiment. It is unclear how the effects in the gender experiment would have shifted if participants were exposed to an equal number of sentences. With that in mind, it is possible to speculate about why this occurred. The gender items contained a baseline of number interference since the grammar of the Portuguese only permits this structure to appear in plural wh-remnants. Therefore, perhaps the combination of having two relevant cues overlapping actually strengthened the effects of the Feature Match X Locality interaction. This idea closely aligns with the Cue (Super) Additivity Hypothesis on the

surface. Unfortunately, sluices do not allow for a concrete way of testing this hypothesis, since both gender and number are relevant features in this construction. As it is currently constructed, the Cue Super Additivity Hypothesis deals with the behavior of a relevant and irrelevant feature, as opposed to two relevant features. We had no initial predictions about the potentially super-additivity that could arise from two relevant features, but if findings like this are found in the future, perhaps an amendment to the current hypothesis should be formulated.

The main take away from these findings, regardless of why gender interacted more strongly with Locality than number, is that both studies show that relevant retrieval features are able to cause retroactive similarity-based interference effects to surface. In addition, we now have evidence that both gender and number serve as retrieval cues in contexts where they are relevant, which is a small but important step towards identifying the possible linguistic information that is involved in retrieval.

4.7 Experiment 3: Gender X Number

Unlike Experiments 1 and 2, which addressed the behavior of two relevant retrieval cues, Experiment 3 was designed to explore the possibilities of the Cue Additivity and Encoding hypotheses. Brazilian Portuguese provides an ideal testing environment for these hypotheses because it allows for both gender and number to be manipulated between the target and distractor, while maintaining number as the only relevant feature for retrieval. This manipulation is possible gender concord is grammatically required but is not relevant for resolving the dependency between the wh-remnant and its correlate.

4.7.1 Design

This experiment used a 2X2X2 design that crossed Locality (Local, NonLocal), Gender (Match, Mismatch) and Number (Match, Mismatch). However, due to the difficulty in recruiting a

sufficient number of subjects for a within-subjects design, this study was conducted using a between-subjects analysis. The two experiments were split by the factor of Locality. Therefore, in one sub-experiment, subjects only saw local sluices containing a 2X2 crossing Gender (Match, Mismatch) and Number (Match, Mismatch), and in the subsequent experiment, a separate group of subjects saw the same manipulation and non-local sluices.

All sentences were identical (modulo the feature manipulations) across both sub-experiments, and as before, the target was always an indefinite NP and the distractor was definite. In order to avoid saturation effects, the 24 experimental sluices in each experiment were interspersed with 12 filler sluices of the opposite target position (e.g. 12 local sluice fillers in the non-local experiment), 5 catch items, 4 ambiguous sluices, 20 unrelated experimental items, and 76 nonexperimental fillers. The sub-experiments were analyzed individually and then combined for the final analysis. The findings of all three sub-experiments are presented below to highlight the trends and differences between them.

4.7.2 Subjects

Subjects participated in the self-paced reading experiments in a computer laboratory at the State University of Campinas in Brazil. A total of 36 subjects participated in the Non-localsluice sub-experiment but only 32 were analyzed. Four subjects were excluded, 2 for self-reporting dyslexia and two for scoring below 80% on filler comprehension question. In the Local Sluice experiment, 34 subjects participated and two were removed for scoring less than 80% on catch comprehension questions. This left a total of 64 subjects in the combined analysis, 32 for each subexperiment.

4.7.3 Analysis

As before, three pairwise anova tests were conducted between the models discussed in the General Analysis section. The best fitting model for each region varied in each sub-experiment. For the

non-local sluices, the baseline model (Number*Gender*Locality) was the best fit for the critical and final regions. The model containing Trial as an interactive predictor was the best fit for the spillover region ($X^2(11)=13.19$, $p<0.01$when compared to baseline). In the local sluicing experiment, the critical and spillover regions showed a trend for Trial as an additive predictor to trend towards a better fit for the data than the baseline models ($p=0.06$ and $p=0.05$, respectively), but this did not reach significance. Finally, when the subexperiments were combined, the baseline model was the best fit for the critical and final regions, and the model containing Trial as an interactive predictor was the best fit for the spillover region ($p<0.05$). Overall, the baseline model was the best fit for the majority of the analyzed regions across experiments, and only the spillover region in all three experiments diverged from this pattern. For this reason and for ease of comparison across the three experiments, the results generated by the baseline model will primarily be reported, but when relevant, I will return to the models of best fit for regions where this pattern diverged.

4.7.4 Predictions

This manipulation has the potential to directly address all four primary hypotheses of this dissertation: Relevant Feature Retrieval, Generalized Feature Retrieval, Cue Additivity and Encoding Interference. Furthermore, the between-subjects design sets the stage for a more theoretical discussion about the differences between proactive interference and encoding interference in local sluices, which could not be discussed in the previous experiments.

The results of this experiment are expected to provide direct support for only one of the retrieval hypotheses (Relevant vs. Generalized Feature Retrieval). If only number matches result in processing delays, either in the form of a main effect or an interaction with Locality, there is direct support for the Relevant Feature Retrieval Hypothesis. However, if both gender and number

show similar effects, the Generalized Feature Retrieval hypothesis would appear to be a more accurate description of the mechanism. Similar effects, in this case, would mean identical main effects of feature match slowdowns and/or similar interactions with Locality between for both the relevant and the irrelevant features. Importantly, these "similar effects" between features should *not* surface as an interaction between Gender and Number, where the Gender Match-Number Match condition is the hardest to process. If this interaction is observed, it would potentially support the Cue (Super) Additivity hypothesis.

The next challenge that needs to be addressed is how we disentangle the Cue (Super)Additivity hypothesis from Encoding interference. As discussed in Chapter 2, encoding interference effects can appear at the retrieval site and be confounded with well-documented retrieval interference effects. Considering that *all* features undergo the encoding process, including those that will be later employed by the retrieval mechanism, it is plausible that a processing delay for Gender Match – Number Match could be a result of both retrieval and encoding interference effects appearing in the same region since Number has undergone two levels of deactivation (see Chapter 2 for mechanism description).

On the other hand, this interaction could easily appear to be a super additive that does not result from encoding interference. Fortunately, from a theoretical standpoint, there is one crucial difference that can separate the two hypotheses. The Cue Additivity hypothesis is retrieval driven. This means that we should see all of the same effects that we would predict under either of the retrieval hypothesis *in addition* to the Gender Match – Number Match penalty. In other words, we should observe the same interactions between Feature Match and Locality, along with main effects of Feature Match that were found in Experiments 1 and 2. On the other hand, encoding interference effects can be separated from retrieval effects because they should not exclusively appear in

retrieval environments (non-local sluices). In the context of this study, this would mean that gender, if it is indeed generating encoding interference, *should not* interact with Locality and should only surface as a main effect regardless of the target position. This would closely reflect the gender effects shown in Chapter 3, where a main effect of gender match was found regardless of clause type.

Taking this one step further and breaking the predictions down by sub-experiment, non-local sluices should show the strongest retrieval interference effects. Given that Number Match was penalized in Experiment 1 in non-local sluices, this effect should also surface in this study. A more interesting realm of theoretical possibilities opens up when considering the predictions for the Local only sluices. Since the retrieval literature predicts more severe interference effects for retroactive interference (non-local sluices), it is unclear how strong the interference effects will be in the proactive case presented in the local sluices, not only because proactive effects tend to be weaker than retroactive ones, but also because local correlates have been the theoretical and statistical baseline cases (see Harris, 2015; Ch. 2 for discussion of proactive vs. retroactive effects). Additionally, the local sluicing analysis may provide insight in separating the encoding effects from retrieval effects. Under the current formulation of the Relevant Feature retrieval hypotheses, regardless of the positional relationship between the target and the distractor, irrelevant features are not expected to generate retrieval interference. Therefore, if the effects in the local sluices are purely proactive retrieval interference, we should still see no effects of gender match, as this feature is not relevant to the dependency in question. Any effect of gender match, as the irrelevant cue, could provide evidence in favor of encoding interference.

4.7.5 Results - Subexperiments

Since the design of this experiment allowed for three separate analyses (non-local sluices, local sluices, and a combination of the two), all three will be reported below. The results are presented in the following order: non-local sluices, local sluices and the overall combined experiment. As before, the final three regions of the sentence and the comprehension questions were analyzed using linear mixed effects regression models in R and the results can be found in Appendix 6, 7, and 8.

4.7.5.1 Non-local Sluices

I will begin by presenting the non-local sluicing results, as this is where the strongest effects were predicted to emerge. The means and standard errors for the non-local sluices are presented below, along with a visual representation of the effects when compared to the grand mean of reading times.

	Critical Region		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	1013 (58)	911 (43)	102
Gender Mismatch	1004 (60)	995 (63)	9
Gender Match Penalty	9	-84	
	Spillover Region		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	939 (39)	872 (37)	67
Gender Mismatch	927 (41)	879 (34)	48
Gender Match Penalty	12	-7	
	Final Region		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	1013 (57)	822 (42)	191
Gender Mismatch	980 (47)	813 (37)	167
Gender Match Penalty	33	9	

Table 29: *Reading Time Results for non-local sluices - means and standard errors in ms*

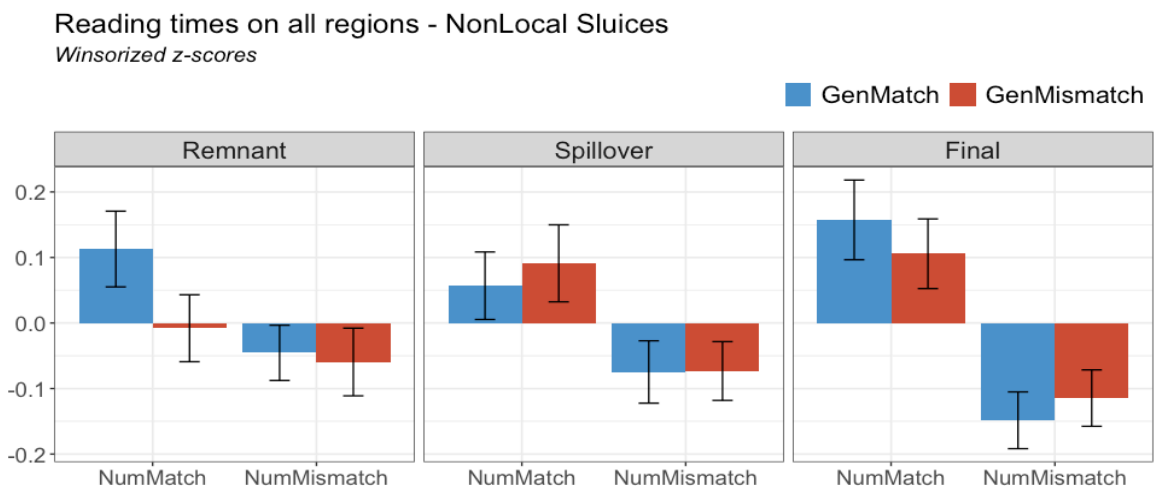


***Figure 20**: Normalized Reading Times across Regions - Non-local sluices*

Critical Region	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	980.580	81.520	31.000	12.028	<0.001
GenderMatch	-18.840	23.520	733.000	-0.801	0.423
NumberMatch	27.700	23.520	733.000	1.178	0.239
GenderMatch:NumberMatch	23.440	23.520	733.000	0.997	0.319
Spillover Region - best fit					
GenderMatch	-18.196	35.739	722.752	-0.509	0.611
NumberMatch	98.486	35.665	721.559	2.761	<0.01
Trial	-1.014	2.526	723.979	-0.401	0.688
GenderMatch:NumberMatch	90.214	35.731	723.607	2.525	<0.05
GenderMatch:Trial	1.519	2.543	727.153	0.597	0.551
NumberMatch:Trial	-5.575	2.537	725.733	-2.198	<0.05
GenderMatch:NumberMatch:Trial	-6.718	2.543	728.340	-2.642	<0.001
Spillover Region* Baseline model					
(Intercept)	904.244	44.165	40.680	20.474	<0.001
GenderMatch	1.341	17.163	710.000	0.078	0.938
NumberMatch	28.600	17.163	710.000	1.666	= 0.096
GenderMatch:NumberMatch	4.828	17.163	710.000	0.281	0.779
Final Region					
(Intercept)	906.809	55.693	37.472	16.282	<0.001
GenderMatch	10.516	20.802	710	0.506	0.613
NumberMatch	89.284	20.802	710	4.292	<0.001
GenderMatch:NumberMatch	5.905	20.802	710	0.284	0.777

***Table 30**: LMER Results - Non-local Sluices*

4.7.5.2 Number Match – Non-local Sluices Only

There was a weak trend for number matches in the spillover region (111 ms, $p=0.096$), which became significant in the final region (358 ms, $p<0.01$). There is also a numerical trend in the means of the spillover region for a Number Match penalty of 115 ms, but this does not reach significance ($p=0.096$ in the baseline mode[14]l). Overall, these findings align with the prediction that relevant features should cause similarity-based interference effects in non-local sluices and the findings in Experiment 1.

4.7.5.3 Gender Match – Non-local Sluices Only

The Non-local sluices did not show any effect of Gender Match. On the surface, this supports the initial predictions made by the Relevant Feature Retrieval hypothesis. Since non-local sluices should show the greatest amount of retroactive retrieval interference, the presence of Number (the relevant feature) interference and absence of Gender interference suggest that the retrieval mechanism is not sensitive to this feature.

4.7.5.4 Interactions in Non-local Sluices

There were no interactions between Gender and Number found for the non-local sluices in the baseline model containing just Gender * Number, as this was the best fit for the majority of the regions across the sub-experiments presented in this section. However, in the non-local sluices, the model containing Trial as an interactive predictor was a significantly better fit for the data in the spillover region. In this model, there was an interaction between Gender Match and Number Match in the spillover region showing that this condition is the hardest to process ($p<0.05$). At face value, this could be considered a super additive effect, but as discussed in the predictions, the lack of gender effects in any other condition suggest that this could be explained by encoding as

[14] This trend is significant in the more complex model treating Trial as an interactive predictor ($p<0.01$)

well. This interaction may be driven by residual effects of encoding interference for both gender and number features since number has undergone both types of interference. The Gender Match – Number Match condition also interacts with Trial, showing that this condition got easier to process over time, which may explain why it does not appear in other regions. If this interaction was weak (albeit significant), the Trial effects may have simply washed it out before it could surface in other analyses.

4.7.5.5 Comprehension Questions

Also reflected in the online findings, Number Match also resulted in less accurate comprehension question responses ($p<0.001$). As a reminder, the comprehension questions asked the subjects to directly identify the target of the sentence, therefore, any incorrect answer is indicative of an incorrect final interpretation of the sentence. Essentially, subjects ignored the semantic infelicity of the distractor (definiteness) in favor of selecting a Local distractor that matched in number with the wh-remnant.

	Comp. Questions		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	55% (4)	84% (3)	-29%
Gender Mismatch	55% (4)	85% (3)	-30%
Gender Match Penalty	0%	-1%	

***Table 31**: % Correct on Comprehension Questions - Non-local Sluices. Means and Standard Errors*

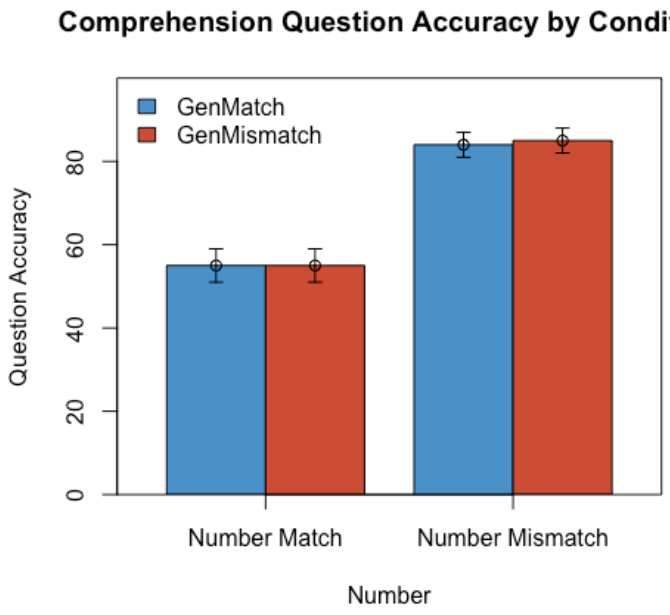


Figure 21: *Comprehension Question Accuracy by Condition - Non-local Sluices*

	Estimate	df	t value	Pr(>|t|)
(Intercept)	1.128	0.205	5.514	<0.001
GenderMatch	-0.034	0.093	-0.369	0.712
NumberMatch	-0.884	0.097	-9.113	<0.001
GenderMatch:NumberMatch	0.035	0.093	0.376	0.707

Table 32: *LMER Results Non-local sluices*

4.7.5.6 Local Sluices Only

This section discusses the results for the Local Sluice sub-experiment. In this section, I assume that any retrieval interference effects are proactive in nature and that any slowdowns caused by non-retrieval features are a result of encoding interference. Although local sluices also require the correlate to be retrieved from the unelided clause, the previous studies on this construction showed that similarity-based interference effects were stronger in the Non-local target position (Harris 2015, 2019). This aligns with the classic retrieval literature, which claims that retroactive interference produces stronger effects than proactive and/or encoding interference. Therefore,

there was a slight prediction that similarity-based interference effects would not surface in local sluices. However, since local sluices have been the theoretical and statistical baseline in the studies and in Harris (2015, 2019), the prediction of the absence of an effect is not particularly strong.

	Critical Region		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	954 (55)	837 (39)	117
Gender Mismatch	796 (36)	733 (31)	63
Gender Match Penalty	158	104	
	Spillover Region		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	876 (34)	812 (30)	64
Gender Mismatch	923 (42)	806 (29)	117
Gender Match Penalty	-47	6	
	Final Region		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	1013 (57)	822 (42)	191
Gender Mismatch	980 (47)	813 (37)	167
Gender Match Penalty	33	9	

***Table 33**: Reading Times Across Regions for local sluices - Means and Standard Errors in ms*

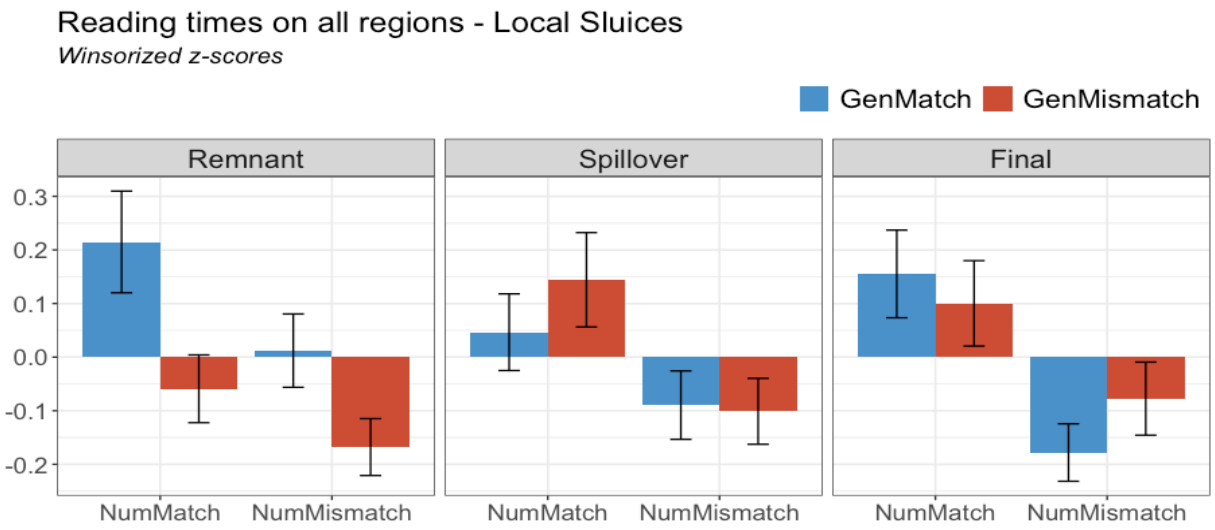


***Figure 22:** Normalized Reading Times Across Regions - Local Sluices*

critical region *simplest	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	830.250	50.960	33.270	16.290	<0.001
GenderMatch	65.480	18.350	710.000	3.569	<0.001
NumberMatch	44.930	18.350	710.000	2.449	<0.05
GenderMatch:NumberMatch	13.570	18.350	710.000	0.739	0.460
critical region *best fit					
(Intercept)	1069.608	108.089	27.426	9.896	<0.001
GenderMatch	117.401	37.849	719.441	3.102	<0.01
NumberMatch	66.281	37.688	709.298	1.759	= 0.079
Trial	-19.142	7.417	28.299	-2.581	<0.05
GenderMatch:NumberMatch	14.244	37.848	719.336	0.376	0.707
GenderMatch:Trial	-4.098	2.655	719.837	-1.544	0.123
NumberMatch:Trial	-1.746	2.640	712.468	-0.661	0.509
GenderMatch:NumberMatch:Trial	-0.016	2.655	719.866	-0.006	0.995
spillover region *simplest					
(Intercept)	854.310	46.910	44.590	18.212	<0.001
GenderMatch	-10.250	14.840	710.000	-0.691	0.490
NumberMatch	45.310	14.840	710.000	3.054	<0.01
GenderMatch:NumberMatch	-12.990	14.840	710.000	-0.875	0.382
spillover region *best fit					
GenderMatch	-74.563	30.995	717.907	-2.406	<0.05
NumberMatch	56.275	31.669	729.329	1.777	= 0.076
Trial	-3.199	2.197	720.818	-1.456	0.146
GenderMatch:NumberMatch	-43.827	31.146	721.080	-1.407	0.160
GenderMatch:Trial	5.097	2.202	721.151	2.315	<0.05
NumberMatch:Trial	-0.815	2.264	734.899	-0.360	0.719
GenderMatch:NumberMatch:Trial	2.445	2.216	725.160	1.103	0.270
final region					
(Intercept)	788.933	44.989	43.669	17.536	<0.001
GenderMatch	-5.435	15.013	710.000	-0.362	0.717
NumberMatch	60.770	15.013	710.000	4.048	<0.001
GenderMatch:NumberMatch	18.491	15.013	710.000	1.232	0.218

Table 34: *LMER Results- Local Sluices*

4.7.5.7 Number Match – Local Sluices Only

Since this study was designed to test the interaction across experiments, it was quite surprising that local sluices showed very strong effects of Number Match in all regions. In addition, the Number

Match penalty got progressively stronger until arriving at the final region (critical region = 180 ms, $p<0.05$; spillover region = 181 ms, $p<0.01$; final region = 358 ms, $p<0.001$).

It was largely unexpected that the Number Match effects would be stronger and more pervasive in the local sluices than in the non-local sluices since this is an example of proactive interference. A potential interpretation of these results is that the stronger effect is a result of encoding and proactive retrieval effects surfacing at the same time, but since the prediction of encoding has not been examined on a similar correlate, this is speculative. However, when proactive retrieval and encoding interference effects are given a chance to surface in a less memory-taxing context, we are able to observe that the effects of similarity-based interference are nearly identical.

4.7.5.8 Gender Match – Local Sluices Only

In this study, the presence of any Gender Match interference effects is of theoretical interest, as it is the best way to identify encoding or super additivity effects in sluiced sentences. The current cue-based retrieval models predict that non-retrieval features will not generate similarity-based interference effects, and for this reason a Gender Match penalty could easily be accounted for by the Encoding Hypothesis. A main effect of gender aligns more closely with the Encoding Hypothesis than the Cue (Super) Additivity hypothesis because the latter does not predict a main effect of gender. Under its current formulation the (Super) Additivity Hypothesis predicts an irrelevant feature matches will only produce an interaction with relevant feature matches, resulting in a greater processing delay when both types of features match. There was indeed a Gender Match penalty in the critical region for local sluices. This is highly reminiscent of the Gender Match effects in the RC region in Chapter 3, where the encoding effects appeared on the critical region and did not persist through the rest of the sentence.

The Gender Match penalty leaves us with an interesting question: Is this an encoding effect or merely a case of proactive interference? Although this experiment cannot completely address this question, it appears that there is stronger evidence for the encoding hypothesis. To begin, let's address the retrieval mechanism. Considering the findings in this chapter and Chapter 3, there appears to be strong support for the Relevant Feature Retrieval hypothesis in cases where the retrieval mechanism is expected to be involved (ORCs and non-local sluices). In both of the constructions tested in this thesis, number either interacts with the syntactic information or exists as a main effect across multiple regions, when it is the only relevant feature in the given dependency. With that in mind, the retrieval mechanism does not show identical effects for gender, which indicates that this feature is likely not involved in retrieval. Therefore, it would be less parsimonious to assume that the retrieval mechanism behaves differently in cases of proactive and retroactive interference. With that being said, if irrelevant features cannot generate retroactive interference, they should not be able to produce proactive interference. Given the lack of relevance of gender in this experiment, I posit that the Gender Match effects result from the encoding mechanism. I will return to this discussion in greater depth in Chapter 5, where I compare the results of both experiments.

4.7.5.9 Interactions in Local Sluices

There were no interactions between Number and Gender in the local sluices.

4.7.5.10 Comprehension Questions

Unlike the non-local sluices, both Gender Match ($p<0.001$) and Number Match ($p=0.08$) influenced comprehension question accuracy. These results are very interesting and rather unexpected, especially in the case of gender match. Since gender is not expressed between the correlate and the remnant, retrieval-only models would predict that this feature should not affect

sentence comprehension. These offline effects point to the involvement of encoding interference as well. As further support for this claim, Villata et al. (2018) pointed out that encoding interference effects are more likely to show up in offline comprehension questions than in online reading times, as the effects are often quite weak online. A potential interpretation of this is that encoding interference degrades the quality of the item in memory, without affecting its accessibility. This approach is largely aligned with the logic of SAT experiments, which show that the slope of an asymptote represents the accessibility of an item, and online effects, and the height of the asymptote represent the quality of the representation, and could result in potential offline effects (see Foraker et al., 2011 for an overview of SAT experiments). It would be interesting to employ a SAT paradigm to studies on encoding effect to see if these predictions are supported.

	Comp. Questions		
	Number Match	**Number Mismatch**	**Number Match Penalty**
Gender Match	86% (3)	93% (2)	-7%
Gender Mismatch	96% (1)	97% (1)	-1%
Gender Match Penalty	-10%	-4%	

Table 35: *% Correct for Comprehension Questions - Local Sluices*

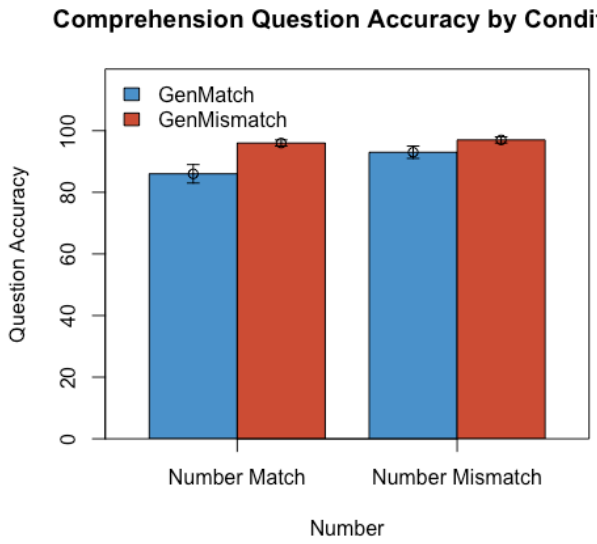


Figure 23: *% Correct for Comprehension Questions - Local Sluices*

	Estimate	df	t value	Pr(>|t|)
(Intercept)	2.966	0.256	11.607	<0.001
GenderMatch	-0.552	0.166	-3.321	<0.001
NumberMatch	-0.287	0.166	-1.729	= 0.084
GenderMatch:NumberMatch	-0.136	0.166	-0.821	0.412

Table 36: *LMER Results - Local Sluice Comprehension Questions*

It should be noted that the overall question accuracy is higher for local sluices than for the non-local sluices, where the lowest percent correct was 55% (discussed in combination experiment). This indicates that subjects were more likely to select the correct correlate when it was in local position, regardless of the feature matches involved.

4.7.6 Results - Non-local + Local Sluices Combined

The ultimate goal of this experiment was to conduct a 2X2X2 experiment in which all of the predictive factors discussed in this chapter are included. Although the results of the sub-experiments may be very insightful, it is not until they are combined that we can address the degree to which the structure of the sluice interacts with relevant and irrelevant features. The table below presented the LMER results and their interpretations are discussed in the following sections.

critical region	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	915.117	60.691	36.494	15.078	<0.001
GenderMatch	23.322	15.557	1466.231	1.499	0.134
NumberMatch	36.319	15.557	1466.231	2.335	<0.05
NonLocal	27.512	17.045	1504.940	1.614	0.107
GenderMatch:NumberMatch	18.505	15.557	1466.231	1.190	0.234
GenderMatch:NonLocal	-42.161	15.557	1466.231	-2.710	<0.01
NumberMatch:NonLocal	-8.615	15.557	1466.231	-0.554	0.580
GenderMatch:NumberMatch:NonLocal	4.939	15.557	1466.231	0.317	0.751
spillover region *baseline					
(Intercept)	873.059	39.804	47.296	21.934	<0.001
GenderMatch	-4.456	11.626	1466.369	-0.383	0.702
NumberMatch	36.955	11.626	1466.369	3.179	<0.01
NonLocal	15.669	12.554	1469.409	1.248	0.212
GenderMatch:NumberMatch	-4.079	11.626	1466.369	-0.351	0.726
GenderMatch:NonLocal	5.797	11.626	1466.369	0.499	0.618
NumberMatch:NonLocal	-8.356	11.626	1466.369	-0.719	0.472
GenderMatch:NumberMatch:NonLocal	8.907	11.626	1466.369	0.766	0.444
spillover region *best fit					
GenderMatch	-45.821	24.032	1469.592	-1.907	= 0.057
NumberMatch	79.130	24.138	1475.687	3.278	<0.01
NonLocal	-3.856	24.399	1493.196	-0.158	0.874
Trial	-2.238	1.701	1470.416	-1.315	0.189
GenderMatch:NumberMatch	26.749	24.090	1472.583	1.110	0.267
GenderMatch:NonLocal	33.847	24.100	1472.436	1.404	0.160
NumberMatch:NonLocal	14.966	24.276	1480.372	0.616	0.538
GenderMatch:Trial	3.259	1.703	1472.500	1.913	= 0.056
NumberMatch:Trial	-3.342	1.714	1480.385	-1.950	= 0.051
NonLocal:Trial	1.481	1.703	1470.945	0.870	0.385
GenderMatch:NumberMatch:NonLocal	58.718	24.113	1473.192	2.435	<0.05
GenderMatch:NumberMatch:Trial	-2.447	1.709	1476.648	-1.432	0.152
GenderMatch:NonLocal:Trial	-2.246	1.711	1476.573	-1.313	0.189
NumberMatch:NonLocal:Trial	-1.866	1.727	1485.976	-1.080	0.280
GenderMatch:NumberMatch:NonLocal:Trial	-3.915	1.711	1477.252	-2.288	<0.05
final region					
(Intercept)	852.404	42.483	45.532	20.065	<0.001
GenderMatch	2.541	13.289	1465.279	0.191	0.848
NumberMatch	75.027	13.289	1465.279	5.646	<0.001
NonLocal	51.060	14.414	1487.008	3.542	<0.001
GenderMatch:NumberMatch	12.198	13.289	1465.279	0.918	0.359
GenderMatch:NonLocal	7.976	13.289	1465.279	0.600	0.548
NumberMatch:NonLocal	14.257	13.289	1465.279	1.073	0.284
GenderMatch:NumberMatch:NonLocal	-6.293	13.289	1465.279	-0.474	0.636

Table 37*: LMER results, Local + Non-Local Sluices*

4.7.6.1 Locality

A main effect of Locality was predicted given the previous findings in this chapter and the literature discussed in the introduction. It was unclear, however, how splitting the experiments on locality to conduct the between-subjects design would affect the strength of this manipulation. To control for this, each sub-experiment (Non-local only and Local only) also contained 12 non-experimental sluices with the opposite target position. In the combined experiment, there was a main effect of Locality in the final region ($p<0.001$). However, there was no interaction between Locality and any other factor, other than the facilitatory interaction discussed above. This was rather unexpected, as Experiment 1 contained penalties for violating the Locality Bias in multiple regions and the sentences were nearly identical to those in this experiment, aside from the manipulation of gender.

4.7.6.2 Number Match – Non-local and Local Sluices

Closely aligning with the Relevant Feature Retrieval hypothesis and the findings in Experiment 1, Number Match was penalized in all regions ($p's<0.05$ for the critical and spillover regions and a $p<0.01$ in the final region). This could potentially be evidence for a sentence wrap-up effect or demonstrate that subjects were waiting for the final region to interpret the sluice. It is unclear, however, if these findings were primarily driven by the local or non-local sluices. The non-local sluices only showed a penalty for Number Match in the critical and final regions, whereas this penalty was present in all regions in the local sluices. It is likely the effect in the spillover region in this combined analysis is driven primarily by the local sluices. Regardless of which sub-experiment contributed the most to these findings, they appear to lend stronger support to the Relevant Feature Retrieval hypothesis than a generalized cue hypothesis, as discussed in 4.7.6.3.

4.7.6.3 Gender Match – Non-local and Local Sluices

The presence of gender-driven interference is not predicted by the Relevant Feature Retrieval hypothesis, but effects of gender match could be indicative of either the Cue (Super)Additivity or Encoding hypotheses, depending on how they surface. If only a main effect is present, this would align more closely with the encoding hypothesis, but if there is evidence of an interaction between number and gender match, we would have stronger support for the super additivity approach. In the combined experiments, there was no effect of Gender Match in any region. There was an interaction in the critical region showing an advantage for the Gender Match – NonLocal condition, but these findings do not help in narrowing down the hypotheses. Although this seems to show that the retrieval mechanism is not sensitive to non-retrieval features, this changes when the experiments are analyzed by Locality, which will be revisited in the discussion section.

4.7.6.4 Interactions

In this study, there were several interactions of interest. To begin, an interaction between Number Match X Locality would support the findings in the Number Experiment (1), as well as the Relevant Feature. Next, an interaction between Gender X Locality would indicate that gender, even when irrelevant to the dependency resolution, is strong enough to modulate similarity-based interference effects, which could support the Generalized Feature Retrieval hypothesis. Finally, a three-way interaction between Gender, Number and Locality showing that when both features match, Non-local targets become harder to retrieval would also support the Generalized Feature Retrieval Hypothesis and potentially the Cue Additivity hypothesis. With all of that said, there were no interactions in any regions in the combined analysis, at least in the baseline model. The lack of interaction lends support to the Relevant Feature Hypothesis, since we do have an across-the-board Number Match Effect in all regions, but this feature might not be strong enough to

modulate the Locality Bias due to the noise generated by using different subject populations in this design.

As a follow-up, I also analyzed the results for the spillover region with its model of best fit, which contained Trial as an interactive predictor ($p<0.01$). In this analysis, the effect structure changed drastically. A main effect of number ($p<0.001$) and a marginal effect of gender match ($p=0.05$) appeared. Both of these factors trended towards an interaction with Trial, such that number matches got easier to processes and gender matches became more challenging throughout the experiment (p's=0.05); however, gender and number did not interact with each other. A three way interaction was identified showing that gender match and number match penalties increased in non-local position, which appears to align with the super additivity hypothesis. Depending on the perspective, these findings could essentially align with any of the hypotheses that consider the behavior of irrelevant features. Looking first at the generalized feature approach, there is evidence that number and gender both generated similar effects in the spillover region. From the super additivity perspective, when both features matched and were presented in non-local position, the sentence became harder to process. Finally, an encoding approach + relevant feature retrieval could potentially explain why the effect of gender was only marginal, when compared to number, if relevant and irrelevant features can both generate interference, albeit from different mechanisms.

Turning out focus towards the interaction once more, sentences containing non-local targets containing both gender and number matches were the hardest to process, and this effect weakened across trials ($p<0.05$). Once again, this could be interpreted either as a super-additive or an encoding effect, where the crucial difference between the two is the degree of interaction that each feature (gender or number) has with Locality. The results of this specific experiment are as to which hypothesis has more support, as neither gender nor number interacted with Locality when

the experiments were combined. When this is considered in conjunction with the RC experiment, it does appear to support the Relevant Feature Retrieval + Encoding approach, as opposed to the Cue Additivity Hypothesis. With the cross-structural evidence that relevant features consistently result in processing slowdowns reminiscent of retrieval interference, as well as the fact that irrelevant features typically appear early on as main effects (regardless of structural information), there is more support for interactions to be interpreted as results of two types of interference.

4.7.6.5 Comprehension Questions

While the online measures did not show any effect of gender in the baseline model, strongly implying that irrelevant cues do not generate similarity-based interference effects, the comprehension question accuracy results showed a different pattern. For the comprehension questions, we observe a main several main effects: Gender Match, Number Match and Locality. Overall, these findings suggest that all feature matches cause comprehension question difficulty, regardless of the relevancy of the cue. Once again, this is corroborated by Villata et al. (2018). who noted that nearly all studies on encoding interference found effects in the offline comprehension questions and not in the online measures.

4.7.6.6 Searching for Encoding Interference in the NPs

Throughout this dissertation, I have claimed that encoding interference could surface either before or at the retrieval site in the form of a main effect of an irrelevant feature. In ideal circumstances, encoding effects would appear at both sites due to the proposed nature of the encoding mechanism. This is because encoding interference should occur regardless of the presence of a retrieval cue, potentially allowing its effects to surface before the retrieval cue is encountered. Although this has been predicted in the literature, to the best of my knowledge, effects of this type have not been previously reported. The sluices allowed us to address this prediction more directly, in a way that

the RCs did not. In all of the sluices, the target and distractor NPs were their own regions, which means that the reading time for NP2s could be measured. Since all features must be encoded, any main effect showing a penalty for either gender or number match can be interpreted as early encoding interference effects.

Before delving into the results, I want to further discuss the prediction that a slowdown will appear on NP2. First, we can revisit the earlier discussion about the feature overwriting mechanism (e.g. Nairne, 1960; Oberauer and Kliegl, 2006). If overwriting occurs when NP2 shares a feature with NP1, making NP1 less active, it seems possible that NP2 might have a longer reading time than NP1. The penalty might occur as activation levels are reassigned between the target and distractor (e.g., feature leveling model in Villata et al., 2018). However, it is important to remember that the feature overwrite model is retroactive (changes to NP1's activation levels only happen after NP2 has been encoded), so it is possible that we would see no effects on NP2 in the given paradigm. The SPR paradigm limits the reading profile to pre-determined chunks and the subject cannot regress to earlier regions, which means that a different pattern could emerge in more sensitive measures than cannot be shown in gross reading times per region.

I conducted post-hoc analyses on the Combined sluicing experiments since these were the only experiments where encoding interference could be observed (gender was relevant in Experiment 2). I employed the same three-model comparison that is reported in the analysis for all experiments, but since I am not trying to compare across regions, the model of best fit for each analysis is relatively unimportant. I predicted that if feature leveling or overwriting actually had an effect on the reading time of NP2, it would result in a slowdown when the features matched and the cognitive mechanism responsible for this process adjusted the activation of NP2. As before, I also analyzed the sub-experiments to observe any patterns.

In the non-local sluices, there was a penalty for Number Match on NP2 ($p<0.05$), but no effect for gender. This finding is encouraging because all features must be encoded at some point. Without some form of encoding interference, there would be no reason for NP2 to show any effect of feature match. The lack of gender effect is slightly unexpected given the expectations from encoding. After this, I tested the local sluices, since this is where the critical region showed a main effect of Gender Match, and once again, there was no penalty for gender or number matches in this region. It is rather unexpected that the local sluices did not show effects for encoding interference, as they show main effects for both gender and number on the critical regions. In addition, when the analyses are combined, the effects that were found in in the sub-experiments no longer surface. In fact, gender match contains a numerical advantage of 96 ms, and the penalty for number match is only 90 ms and not significant.

As was mentioned at the beginning of this section, I do not consider the lack of effect on NP2 to be concrete evidence that encoding interference has not taken place. The presence of main effects associated with irrelevant features, although they could reasonably be explained by something like a Generalized Feature Retrieval model, appear to align more closely with the predictions for encoding interference made in the literature (e.g., Jäger et al., 2015; Villata et al., 2018). Even though the presence of a slowdown on NP2 would have provided stronger evidence in favor of encoding interference, to the best of my knowledge, reading time effects associated with encoding interference before the retrieval site have only been found in a few studies (Acheson and MacDonald, 2011; Kush et al., 2015 in Villata et al., 2018).

The lack of effects might largely be due to the experimental paradigm being employed. Although self-paced reading is a relatively standard method to examine reading time effects in psycholinguistics, it might not be sensitive enough to capture fine-grained effects of encoding

interference. Previous research by Villata et al. (2018) suggests that encoding interference effects tend to be quite weak, and SPR certainly is not the ideal method for identifying weak effects. Although some research defends the validity of self-paced reading (e.g., Mitchell et al., 2004), it is possible that other more sensitive measures, such as eye-tracking, would be able to find patterns of slowdowns or other reading difficulties (e.g., more regressions in or out of the region) representative of encoding interference. It is impossible in the studies discussed in this thesis studies to determine the exact point where the slowdowns occur since there is no way to look at each individual word or constituent. It would have been extremely convincing to find effects on NP2, but their effects may be dependent on the method or strategies afforded by different methods.

4.7.7 Discussion of Experiment 3

In this set of experiments, we were ultimately looking for one of two interactions: either an interaction between Number and Locality, where non-local targets containing a number match incurred greater reading times or a three-way interaction between Number, Gender and Locality, which would show that when both features matched and the target was NonLocal, reading times would be slower. Finding the first interaction would support Harris' (2015, 2019) findings that the amount of similarity-based interference caused by number cues is modulated by the position of the target, as well as support the current claims in the literature that only feature matches between a retrieval cue and its target/distractor will cause enough competition to generate similarity-based interference effects. If the second interaction were found, or even just a main effect of gender match, we would potentially be able to make claims about multiple sources of processing slowdowns, such as cue additivity or encoding interference. With that being said, neither interaction appeared in the 2X2X2 analysis. Instead, a general penalty for number match and non-local targets appeared. We interpret these results as support for a Locality Bias and for

the Structurally Driven Retrieval Hypothesis, which has not been previously reported for Brazilian Portuguese, as well as similarity-based interference caused by number matches.

Taken at face value, it appears that this final experiment would put to rest any discussion of encoding interference as a source of processing delays, but this is not the final story. When examining the sub-experiments, a main effect of Gender Match appeared in the critical region containing the wh-remnant. This finding is identical to what was shown in the Gender X Number manipulation for Relative Clauses, in which gender was also irrelevant to the grammatical dependency. The fact that gender matches are able to cause processing delays when they are not relevant to the dependency being processed is compatible with an encoding approach. As further support for this claim, all of the experiments conducted throughout this dissertation have provided suggestive compatible with a retrieval mechanism that is sensitive to relevant features, which then interact and cause similarity-based interference effects.

4.8 Conclusion

In conclusion, the results of all three sets of experiments on Brazilian Portuguese sluices were highly consistent with the current models of cue-based retrieval, but also provide support for encoding interference in a way that is not accounted for in the current literature. Experiments 1 and 2 were conducted to gather support for the Relevant Feature Retrieval and Structurally Driven Retrieval Hypotheses and were successful in this endeavor. Although the results varied slightly, both studies showed that relevant retrieval features cause similarity-based interference and that the Locality Bias influences Brazilian Portuguese sentence processing. The expected interaction between Locality and Feature Match only appeared in the subset of the gender sluices discussed in Section 4.6.6.3, but a very strong numerical trend for this interaction appeared in the final region of Experiment 1. Overall, these findings not only support the concept of cue-based retrieval but

also help to narrow down the open question about what type of linguistic information is a retrieval cue. Taken in conjunction with Chapter 3, I propose that a retrieval cue is any linguistic feature that is required to resolve a grammatical dependency.

More interesting than corroborating current theory, however, is that the set of experiments in this chapter provides evidence of encoding interference in a structure that has not been analyzed for these effects. Being able to make a direct comparison about the role of gender morphology when it is irrelevant to the grammatical dependency across two very distinct structures (RCs and Sluices) supports the idea that encoding is always happening, and just like retrieval, the encoding system is sentive to feature overlap and competition. Given the effects of irrelevant features found in distinct structures, methods, and languages, encoding interference and the predictions of feature overwrite warrant future research.

Chapter 5: Conclusions

5.1 Overview

Before delving into the discussion about the relationship between the experiments conducted in this dissertation and the proposed hypotheses, I would like to present the main findings from each study as a point of reference for upcoming discussions.

5.1.1 Relative Clauses

The relative clause experiments were conducted for several reasons. First, I wanted to corroborate the claims that similarity-based interference is a crucial factor in the strength of the relative clause asymmetry that has been discussed so frequently in the psycholinguistic literature. More specifically, I aimed to replicate the findings from Gordon (2001) showing that that interfering information makes the ORC penalty significantly stronger. This replication study is important, as it would support some of the claims made about cue weighting systems prioritizing structural information. If structural cues are more highly weighted, this would account for the ORC penalty ([+Nominative] overlap in ORCs and not SRCs) and the fact all other cues should only generate interference effects when structural features initially match.

To address these topics, I manipulated the number feature that is required for subject-verb dependency resolution in Portuguese under the assumption that this relevant feature would cause similar effects to those observed in Gordon (2001). In addition to these goals, I also wanted to expand on the claims made about the nature of encoding interference in sentence processing (Villata et al., 2018). To accomplish this, I also manipulated gender features, which are grammatically required but not involved in the RC dependency. Below, I present a table of the main findings.

	ORC Penalty	Feature Match Penalty	Interactions of interest
(1) Number X Clause Type	• RC Region	No effect	• Number Match, ORC (trend)
(2) Gender X Clause Type	• RC Region	• RC Region	• Gen Match - ORC facilitation at post-RC region • Gen Match - ORC penalty at matrix verb spillover
(3) Num X Gen ORCS	N/A	• Gender Match, RC	• Gender Match, Number Mismatch is the hardest

Table 38: *Overview of all RC results*

Without delving into the hypotheses, these results demonstrate a few important patterns. First, Brazilian Portuguese speakers are sensitive to the ORC/SRC asymmetry, which to the best of my knowledge, has only been shown in one other study (Gouveau, 2003). Not only does this provide important cross-linguistic information about the behavior of RCs in sentence processing, but it also serves as a sanity check for the functionality of the experiment. Given the vast cross-linguistic evidence for object-extraction penalties, if this effect had not been found, it would be more likely the result of an experimental design flaw than an actual representation of sentence processing in Portuguese. Something surprising about these results, when compared to the previous literature on the topic, is that the ORC penalty appears at the RC region itself, as opposed to the matrix verb. The majority of studies on this topic predict an effect on the matrix verb, as this is the site of retrieval (e.g., Staub, 2010).

Furthermore, Experiment 1 provides clear evidence that number features are strong enough retrieval cues that they can modulate the strength of the RC asymmetry. This supports the claims made by Gordon (2001) that the memory system is one of the driving forces behind object-extraction penalties. It is rather curious, however, that there was no main effect for number match in Experiment 3, where gender was also manipulated. I will discuss potential reasons behind the absence of this effect in upcoming sections.

5.1.2 Sluices

Much like the relative clause experiments, the sluices were also intended to replicate previous findings (Harris 2015, 2019), as well as provide insights into the topics of cue combinatorics and encoding interference. Sluices provide an ideal testing ground for these topics for two main reasons. First, Brazilian Portuguese allows for optional gender marking on wh-remnants via the addition of a pronominal element, which allows (1) a direct comparison between number and gender morphology when they are both relevant for retrieval and (2) gender to become an irrelevant feature for the dependency when the pronominal element is removed. Second, although correlate-remnant pairing in ellipsis requires a pairing mechanism akin to retrieval, very few studies have explored this topic in relation to memory and interference. To the best of my knowledge, only two studies on the relationship between retrieval and sluices have been conducted on English (Harris 2015, 2019), and these studies served as an inspiration for the design of experiments in Chapter 4.

In line with the previous literature, I predicted a penalty for NonLocal sluices, a potential main effect of feature match (when the feature was relevant), and an interaction showing that Feature Matched, NonLocal sluices were the hardest to process. Like Chapter 3, I took these predictions one step further and tested for potential effects of encoding interference in Experiment 3, where number was required in the dependency and gender was not. Below, I present the main findings across experiments.

	NonLocal Penalty	Feature Match Penalty	Interactions of interest
(1) Number Sluices	• Critical and Spillover Regions • Comprehension Questions	• Critical and final region • Comprehension Questions	No interaction
(2) Gender Sluices	No main effect	• Critical Region • Spillover Region • Final Region • Comprehension Questions	• Penalty for NonLocal – Gender Match Final Region
(3) Number + Gender Sluices	• Final Region	• Number Match: all regions and comprehension questions	• No 3-way interaction showing increased difficulty for Gender and Number Matched NonLocal sluices in simplest model *Shown in complex model
(3.1) NonLocal	N/A	• Number Match: Final region	• No Gender Match – Number Match penalty in simple model *Gen X Num appears in complex models
(3.2) Local	N/A	• Number Match: all regions, comprehension questions • Gender Match: critical region, comprehension questions	No interaction

Table 39: *Overview of all sluicing results*

Without discussing their significance to the hypotheses, I will once again highlight the key patterns observed across the experiments. To begin, Brazilian Portuguese speakers are indeed sensitive to the Locality Bias in sluiced constructions, corroborating Harris' (2015, 2019) findings on English. This is shown through the main effects of relevant features in Experiments 1 and 3, as well as interactions between Feature Match – Locality in Experiment 2, and as well as the 3-way interaction that arises in Experiment 3. Moreover, this study demonstrates that when both gender and number are relevant for retrieval, the generate processing slowdowns that have been found in a variety of studies relating to retrieval and similarity-based interference. Finally, when gender is not relevant to the dependency (Experiment 3), it still generates interference effects that are not predicted by most cue-based retrieval models.

5.2 Addressing the Hypotheses

In Chapter 2, this dissertation presented four hypotheses that discuss how the retrieval mechanism will treat relevant and irrelevant features in online processing. Three of the hypotheses were focused on the retrieval mechanism, and the final hypothesis addressed the encoding mechanism. The three retrieval hypotheses generated different predicted surrounding the behavior of irrelevant cues in retrieval. The strongest hypothesis was that of Relevant Feature Retrieval, where only features specifically required by the dependency were expected to produce interference effects. The Cue (Super)Additivity hypothesis considered a greater variety of features, proposing that retrieval is primarily driven by relevant features but that irrelevant features will result in super-additive effects at the retrieval site. Finally, the Generalized Feature Hypothesis that any overlapping feature would be considered by the retrieval mechanism, regardless of its relevancy. The Generalized Feature Hypothesis acts as the null hypothesis in the set of retrieval hypotheses, as it is the one that predicts that feature relevancy should not affect the behavioral profiles or strength of interference.

In contrast with the retrieval approaches, the Encoding Interference Hypothesis was put forth as an alternate explanation for any interference arising from irrelevant features. The primary difference between the retrieval approaches and the encoding hypothesis is that the features involved in encoding should not be sensitive to syntactic information like clause type, resulting in main effects instead of interactions. Another difference is that encoding effects might appear earlier than the retrieval site since this process happens before retrieval begins. One of the challenges with encoding interference, from an empirical standpoint, is that its effects can be conceptually confounded with retrieval interference. This is because encoding interference degrades the representation of items stored in memory and ultimately decreases their retrievability.

Following the evidence in the literature regarding the role of syntactic information, the variability of cue behavior across structures, and the few studies that have also claimed that encoding interference can result from gender in Romance relative clauses (see Chapter 1 for an overview), my underlying assumption was that these studies would support a combination of two hypotheses. In particular, I expected to find evidence for a combination of the Relevant Feature Retrieval and Encoding Hypotheses. Although my interpretation of the results closely aligns with this assumption, I also present a section on the other two hypotheses and leave the ultimate conclusions about the mechanisms involved in interference an open for future research.

5.2.1 The Relevant Feature Retrieval and Encoding Hypotheses

The existing literature on retrieval and encoding, as well as the findings in this dissertation, point in the direction of two or the four hypotheses: The Relevant Feature Retrieval and the Encoding Interference. By combining these two hypotheses, we can deduce that no *irrelevant* feature (those not required to resolve the dependency) should ever cause retrieval-driven similarity-based interference effects. If these features were to cause interference, it is more likely that they derive from the encoding mechanism than the retrieval mechanism. These speculations are line with claims made by several other researchers, who also argue that processing disruptions caused by irrelevant features as encoding interference (Jäger, 2015, Van Dyke and McElree, 2006; Villata et al., 2018).

The predictions, then, made by these hypotheses are quite clear. Under the strongest interpretation of the Relevant Feature Hypothesis, number should always be a source of similarity-based interference in Brazilian Portuguese due to its ability to identify the target (from among other competitors) in both subject-verb dependencies and correlate-remnant pairing. The only experiment where number was not predicted to affect the results was in the Gender Sluicing

Experiment (2), where gender was the relevant feature being manipulated. It is also important to note that both gender and number should produce similar effects when they are the relevant feature to the dependency, as this hypothesis assumes that all relevant features will behave in the same manner. In all other experiments in which gender was not a relevant feature, the strictest interpretation of this hypothesis would predict *no interference effects* for gender cues, at least at the retrieval site. The location of these effects is crucial, as the majority of cue-based retrieval models would predict that retrieval-based interference will occur at the retrieval site itself (the matrix verb for the RCs and the wh-remnant for sluicing) (e.g. Van Dyke and McElree, 2006).

5.2.1.1 Sluicing Evidence

The support for these two hypotheses varied between the RC and sluicing results. To begin, the sluicing experiments provided a significant amount of support that both of these hypotheses operate in tandem, perhaps with a few minor modifications to the Relevant Feature Hypothesis. The relevant feature in Experiments 1 and 3 was number. In Experiment 1, number match resulted in slower reading times on the critical and final regions, as well as decreased comprehension question accuracy. Although it did not interact with Locality, these findings strongly suggest that number was a retrieval cue in this experiment. Experiment 3 patterned similarly, containing a match penalty in all regions, decreased comprehension question accuracy, and no interaction with Locality. These patterns were also reflected in the sub-studies of Experiment 3, which were split on the factor Locality for sub-analyses.

More noteworthy was the effect of gender features in the sluices. In Experiment 2, gender was the only morphosyntactic feature that could be used to distinguish the target from its competitors. As predicted by the Relevant Feature Hypothesis, gender matches incurred a processing delay in all regions, and these slowdowns significantly worsened in NonLocal sluices

in the final region. Looking at the behavior of both gender and number, when relevant to the dependency, the results are highly comparable. The primary difference between the results of Experiments 1 and 2 is the presence of the interaction between Gender and Locality in Experiment 2. Although the relevant feature hypothesis predicts that these cues should behave identically, the interaction in Experiment 2 may be a result of the retrieval mechanism dealing with interference from two sources. Support for this comes from the fact that Experiment 1 showed a strong numerical trend towards this interaction, but perhaps, this feature is not weighted strongly enough to produce the intended interaction on its own, assuming that the retrieval mechanism assigns inherent weights to features. However, when both gender and number match, these cues are strong enough to interact with Locality. Unfortunately, there is no way to completely disentangle gender and number matches in this construction, where both would be relevant to the dependency.

Shifting away from the retrieval mechanism, I will now discuss the behavior of irrelevant gender features in Portuguese sluices. As discussed throughout this dissertation, encoding and retrieval effects tend to be challenging to separate from one another. Ideally, encoding effects would appear before the retrieval site. One prediction is that a slowdown would be found on NP2 in the sluices, as this feature would generate encoding interference with NP1, but NP2 effects have been found in very few studies (Villata et al., 2018 provide a discussion). In addition, it was unclear if the self-paced reading paradigm would be sensitive enough to find these effects, since they provide a relatively limited reading profile. The results in the sluicing experiments did not show any patterns for slowdowns on NP2 in any experiment when gender features overlapped. A second prediction, and more aligned with the current studies, is that encoding effects might surface at retrieval sites as a direct result of the quality of information stored in memory being diminished before retrieval. (Van Dyke and McElree, 2006). In other words, interference effects from

irrelevant features can appear at the retrieval site alongside retrieval interference. The studies in this dissertation loosely supported this claim, as there were main effects of gender match occurring in tandem with retrieval effects.

Looking at the effects of gender that are challenging to explain in a retrieval-only approach, experiment (3) contained an offline effect of gender match in the comprehension questions and an online effect of gender match at the wh-remnant and in the comprehension questions for the Local sluices analyzed in Experiment 3. Although this could be a result of proactive retrieval interference, the Relevant Feature Hypothesis, in isolation, would predict that gender would not modulate interference due to its status as an irrelevant feature. Finally, although it is unclear if these findings directly support a retrieval or encoding approach, there is a three-way interaction in the post-RC region showing that Number and Gender Matched NonLocal sluices were the hardest to process. If encoding effects are strong enough to degrade retrieval accuracy, the involvement of gender in this interaction could logically be a result of encoding interference surfacing alongside standard retrieval interference.

5.2.1.2 Relative Clause Evidence

While the sluicing studies provide support for a combination of the Relevant Feature Retrieval and Encoding interference hypotheses, the relative clause results are less clear. To begin, although number was the only relevant feature in all of the RC studies, it did not generate many results that could be interpreted as similarity-based interference. In Experiment 1, there was a marginal trend showing that ORCs with number matches were the hardest condition to process. This effect is reassuring, as it shows that number is a strong enough cue to modulate the RC asymmetry, at least to some degree. What is concerning, nevertheless, is that this same effect did not appear in Experiment 3, which also contained a gender manipulation.

Although the weak effects of number do not immediately challenge the Relevant Feature Retrieval hypothesis, the behavior of gender in the RC studies raised a few questions. To begin, both Experiment 2, where gender was manipulated and number matched across conditions, and Experiment 3, where the two features were crossed, showed main effects of gender at the RC region. Initially, these effects seem to support the predictions of the Encoding Interference Hypothesis. In fact, they are very similar to what was observed for the Local sluices and comprehension questions for sluicing Experiment 3. Where this interpretation becomes more complicated, however, is when the interactions between Gender and other factors are considered.

In RC Experiment 2 (gender manipulation), there are two contradictory interactions. First, in the RC-spillover region, there is a facilitation effect for gender matches in ORCs. This interaction is quite odd, as none of the hypotheses would predict that a feature match would facilitate processing. Although some literature on cue-based retrieval, in particular, agreement attraction errors (e.g. Wagers et al., 2009) show facilitatory interference effects, it is unclear why gender should interact with clause type at all. Facilitatory interference is typically claimed to be the result of partial matches between a retrieval cue and an encoded item *sometimes* accepted as grammatical, forming an illusion of grammaticality (Parker et al., 2017), but this line of argumentation cannot be applied to the RCs in Chapter 3. All of the RCs were completely grammatical, therefore, there should be no illusions of grammaticality present.

Even more challenging to explain is that the facilitatory interaction flips at the matrix verb spillover region, which seems to contradict the predictions of the Relevant Feature Hypothesis. The initial predictions for this hypothesis were that retrieval interference will arise most strongly when relevant features interact with structural information, particularly when the distractor follows the target. Therefore, the strictest interpretation of this hypothesis would assert that *any* interaction

between a morphosyntactic feature and structure is evidence for retrieval interference. Nevertheless, it is the lack of involvement of irrelevant features that specifically separates the Relevant Feature Hypothesis from a more generalized type of retrieval in which all features are considered. The interaction between Gender Match and ORC in the RC studies is less complex, in regards to the hypotheses, than it initially appears. Although Experiment 1 only produced a marginal interaction between Number Match and Clause Type, these findings still demonstrate that number generates interference in ORCs. In Experiment 2, all of the conditions were matched for number, and gender was manipulated independently. Therefore, the interaction that surfaced at the Matrix Verb could be explained by a combination of retrieval interference and encoding interference, where the former is driven by number and the latter by gender. As in the sluices, support for this claim comes from the fact that gender matches generate early main effects that do not interact with clause type (encoding interference) and number matches do not.

Although Experiments 1 and 2 can be interpreted as support for a combination of retrieval and encoding interference, Experiment 3 was much harder to explain. First and foremost, there were no penalties for number match in any region. Given that all of the sentences were ORCs and number was the only relevant feature, the lack of number interference is rather disconcerting. Also challenging to align with the hypotheses was an interaction showing that the hardest condition was Gender Match – Number Mismatch. These findings appear to directly contradict any of the retrieval hypotheses presented in Chapter 2, and for now, it is unclear as to why this happened. As discussed in Chapter 4 and below, this could have been a result of language-specific cue hierarchies or even an experimental design flaw. Overall, the sluicing and relative clause studies support a Relevant Feature + Encoding approach, more than the other hypotheses (addressed in section 5.3).

5.2.2 What happened with number in the relative clauses?

As the discussion above suggests, the sluicing experiments and the number experiments generated very different results. In fact, the findings varied so much that it became challenging to find concrete support for any of the four hypotheses discussed in this dissertation. In this section, I will speculate as to why the RC experiment did not produce the results that were expected.

5.2.2.1 Experimental Design

First and foremost, I would like to present the most obvious reason that the RCs did not show the same amount of interference effects as sluices: experimental design. To begin, the studied that mine was modeled after (Gordon et al., 2001) employed a word-by-word self-paced reading paradigm. There are pros and cons to this method of presentation. While it allows for effects on each word to be analyzed, it forces subjects to spend time on elements that are oftentimes skipped in more naturalistic reading measures, such as eye-tracking, thereby decreasing its naturalness. I selected to use syntactic constituent chunks in the SPR study primarily to keep the reading process as natural as possible, and to make sure that the RC items and sluicing items were presented similarly. With that being said, I could not test for effects on every dependency site (RC and Matrix verb), nor could I test for encoding interference effects on either NP.

In addition to the chunking of the RC items, each experiment was presented in conjunction with roughly 110 other sentences, including a sluicing experiment. After each participant finished the study, I conducted a brief interview with them for screening purposes. One of the questions asked to all participants was "What patterns did you observe?" In response, the vast majority of subjects made a comment related to the sluicing sentences. Although there was a sufficient number of distractors to evenly space out the sluicing and RC experiments, it appears that subjects were more aware of the manipulations in the sluices than the RCs. If subjects had developed a strategy

to only pay attention to the sluices for grammatical information, since they would be asked directly about the remnant-correlate pairing, it is possible that they paid less attention to all other sentences in the study. Since the questions following the RCs and other fillers asked about the general context of the sentences, it could have been the case that they simply read these sentences for the "gist" and paid less attention to the features that would determine the final parse (e.g., "Good Enough Processing," Ferreira et al., 2002). There is no direct way to verify this claim, but it could be used to explain why the RC results were much weaker than the sluices. If this were truly the case, then perhaps the RC results do not imply as much about the retrieval and encoding mechanisms as they do about experimental learning strategies.

5.2.2.2 Cue Weighting and Cue Diagnosticity

Moving away from the non-theoretical approach of experimental design to explain the RC behavior, I return to the discussion that is spelled out in Chapter 4: cue-weighting. As I proposed in Chapter 4, all of these studies were conducted on a dialect of Brazilian Portuguese that accepts non-standard number agreement for subject-verb dependencies and DP-internal agreement processes. I was aware that the dialect being studied contained this pattern of non-standard agreement. My initial intuition was that it would not have a large effect on reading studies, as non-standard agreement is highly stigmatized in written text and all of the sentences contained standard agreement. Moreover, I specifically included a few (2-3) catch items with improper subject-verb agreement to make sure that subjects were paying attention, and all subjects included in the analysis scored higher than 80%. However, this does not mean that this aspect of their dialect does not affect how strongly number cues are weighted. If cue-weighting can be affected by linguistic preferences, this is an important concept to explore in future studies. I propose that this could be tested on other dialects of Portuguese, initially, to see if that affects overall sentence processing.

If other dialects showed stronger effects of number than the given dialect, it would be strong evidence that cue weighting is much more fine-grained than syntax versus semantics, as models like Van Dyke and McElree (2011) have presented.

This idea of cue-weighting can also be used to explain why number effects were stronger in the sluices. To the best of my knowledge, non-standard agreement cannot be used in the wh-remnant-correlate agreement. According to my non-native intuitions, it does not appear that number can be eliminated from either the target or the dependency without creating an ungrammatical sentence.

(1) a. Os professores avaliaram algumas conclusões, mas eu não sei quais.
The$_{\text{-MPL.}}$ professors$_{\text{M.PL.}}$ evaluated some$_{\text{F.PL.}}$ conclusions$_{\text{M.PL.}}$, but I don't know which$_{\text{PL.}}$

*b. Os professores avaliaram alguma conclusão, mas eu não sei quais.
The$_{\text{-MPL.}}$ professors$_{\text{M.PL.}}$ evaluated some$_{\text{F.SG.}}$ conclusions$_{\text{M.SG.}}$, but I don't know which$_{\text{PL.}}$

*c. Os professores avaliaram algumas conclusões, mas eu não sei qual.
The$_{\text{-M.PL.}}$ professors$_{\text{M.PL.}}$ evaluated some$_{\text{F.PL.}}$ conclusions$_{\text{M.PL.}}$, but I don't know which$_{\text{SG..}}$

Therefore, the definition of dialectal cue-weighting also needs to consider the construction that is being analyzed. Overall, creating a cue-weighting model that can predict the strength of a given cue in different constructions is quite challenging, but not impossible.

I would also like to briefly highlight how this approach aligns with claims regarding cue diagnosticity (Harris, 2015, 2019; Martin, 2016). Several authors have made the argument that cues vary in their diagnosticity and that highly diagnostic cues are better at selecting a target from its competitors and generate distinct interference profiles (Harris, 2015, 2019). Empirically, Harris (2015, 2019) showed that the time course of Locality effect was modulated by the diagnosticisty of the remnant. This approach raises questions regarding the relative clause effects presented above since number should have been a highly diagnostic cue according to the grammar of Portuguese. This assumption would generate the prediction that number would produce earlier effects of

interference than an irrelevant cue like gender, but this was not the case. Number effects appeared sporadically throughout the sentence or not at all. Even if we were to treat number as a weakly diagnostic cue due to its low probability of surfacing in this dialect (e.g., Martin, 2016), we should have seen interference effects at later regions of the sentence, but no patterns to this effect were found in the data. In fact, gender effects appeared more consistently and earlier in the sentence than number effects. This effect was unexpected because gender is an irrelevant feature and would not be expected to be highly diagnostic under most retrieval models. The relationship between cue relevancy and diagnosticity is interesting topic to investigate, but for now, I do not feel that the effects between relevant and irrelevant should be directly compared.

Since the types of dependencies being explored are very different, and the probability of the number feature surfacing really only influences the RCs, more research is needed to explore the relationship between diagnosticity, cue relevancy and the overall strength/probability of a cue. For example, when gender and number were both relevant, which should make them highly diagnostic, they both presented similar temporal profiles for interference (sluicing studies). This finding, when compared with the RCs, could imply 1) that the relevancy of a cue modulates its diagnosticity and/or 2) the diagnosticity of a cue varies structure to structure. Although I do not intend to delve into this discussion, my brief attempt to reconcile unique patterns in this dissertation is the following: cue diagnosticity is not the same process as cue weighting, and diagnosticity is highly correlated with relevancy. Perhaps cue-weighting procedures are experiential in nature and depend on the likelihood of certain cues being available during retrieval, whereas diagnosticity is driven by grammatical factors. When these two factors come into competition, such as the case of a highly diagnostic but low weighted feature like number in Portuguese RCs, interference effects are less predictable. At this time, I am unable to provide further reasoning about how structural

cue weighting and diagnosticity are unrelated enough that they would produce different predictions, but I would like to present the idea that cue-weighting is experientially driven (which can vary by dialect), whereas cue-diagnosticity takes into consideration all of the linguistic information that contributes to the probability of a given cue (or combination of cues) being used in a dependency. This slightly contrasts with claims that cue weighting tends to assign greater weights to syntactic features, but these two approaches can be reconciled with further exploration (e.g., Parker et al., 2017). In sum, a weakly weighted syntactic cue might not necessarily be weakly diagnostic, and vice versa.

5.3 Revisiting the Generalized Feature Retrieval and Cue-Additivity Hypothesis

Although the evidence in support of the Relevant Feature Retrieval and Encoding hypotheses is stronger, I want to address the remaining two hypotheses that were presented in Chapter 2. The Generalized Feature Retrieval Hypothesis claims that the retrieval mechanism is sensitive to all features that overlap between items in memory, regardless of their relevancy to their involvement in the grammatical dependency. Unlike the relevant feature approach, where no interference is expected from features that are not involved in the dependency resolution, this hypothesis proposes that the retrieval mechanism is broadly sensitive to all linguistic features. This would generate the prediction that we would observe no difference between gender and number in any experiment, regardless of how relevant they are for grammatical dependency resolution.

5.3.1 Generalized Feature Retrieval

The Generalized and Relevant Feature Hypotheses partially overlap in their predictions. As in the Relevant Feature Hypothesis, all effects of number when it is relevant are also predicted by this hypothesis. The only way to distinguish between these two is to look at the behavior of gender features when they are not relevant to the dependency. In its strongest form, this hypothesis would

predict that the interference effects for both relevant and irrelevant features will be identical. Across the experiments, there is only one effect that aligns with this prediction. In the RC gender manipulation, gender was not expected to have any effect on processing in the form of an interaction between Gender and Clause Type. However, this interaction was observed in the matrix verb spillover region. In addition, the combined sluicing experiment (Experiment 3) demonstrates a similar effect in the post-RC region. There is an interaction between Gender, Number, and Locality, indicating that gender matches also influenced the processing of these sentences when a structural bias was violated. Nevertheless, it remains unclear if these are truly related to retrieval. As mentioned in the introduction, encoding effects can also appear at the retrieval site in conjunction with retrieval interference. Therefore, it unclear if gender was genuinely being treated as a retrieval cue or if the effects of encoding interference were simply additive when retrieval interference was already present.

As evidence against this hypothesis, gender matches were consistently penalized in cases where no retrieval interference was expected (SRCs, Local sluices). Since this hypothesis predicts that *all* features will produce identical effects if they are involved in retrieval, the varying behavior of number and gender match penalties does not support this prediction. Returning to the concept of cue-weighting, however, perhaps predicting identical behaviors for all features is too strong. If number is a weaker cue than gender in this dialect of Portuguese, it is possible that these effects could be accounted for using this hypothesis. Although I cannot definitively conclude that retrieval was not involved in these effects, I rely on support from other scholars studying encoding interference. Several authors have formulated experiments similar to mine in languages like German, Swedish and Italian, where gender was not a relevant feature to the dependency, under the assumption that this feature would not be involved in retrieval (e.g., Villata et al, 2018; Jäger

et al., 2015, 2017). Given the support in the literature, the effects of gender observed in this dissertation align more closely with previous claims encoding effects than with purely retrieval interference, although both approaches still warrant future consideration.

5.3.2 Cue Additivity

The Cue Additivity hypothesis claims that when relevant features overlap, irrelevant feature matches will cause super-additive processing penalties. Crucially, this means that irrelevant feature overlap should not result in retrieval interference on their own. The intent behind this approach was to account for any effects of number and gender match observed in the RC and Sluices Number X Gender experiments. Breaking down the experiments, there were differences in additivity effects in sluices and RCs. In the sluicing study, there was a super-additive effect that surfaced in Experiment 3, showing that Gender Matched - Number Matched - NonLocal sluices were the hardest to process. This interaction between gender and number also appears in the NonLocal sluices that analyzed as part of Experiment 3. However, as I have explained, these effects align more closely with an encoding approach than an additivity effect. This is because gender was able to produce similarity-based interference on its own, and this hypothesis predicts that it will only surface when relevant features also matched.

The RC findings provide more evidence against this hypothesis. Instead of finding a super-additive effect of feature match in the RC combined experiment (3), a tradeoff was observed. The tradeoff showed that Gender Matches generated a processing slowdown when number mismatched; in fact, at the RC itself, the condition where both features matched had the fastest reading times. As always, it is possible that the weak number penalties in RC Experiment 3 were driven by factors such as cue-weighting. Yet, when the results are taken at face value, especially

the fact that condition where both features matched was facilitated in the RC itself, this provides strong evidence against the Cue Additivity hypothesis as it is currently formulated.

5.4 Addressing the Theory of Cues and Retrieval

I would like to return to one of the initial research questions presented in the introduction of this thesis: What is a cue? Although cue-based retrieval models are beginning to be adopted as the primary model of the memory-sentence processing interface, a fair amount of research is still required before the mechanism can be genuinely understood.

As Parker et al. (2017) point out, to date, there is no fully accepted theory of retrieval cues. This means that it is unclear what linguistic information is actually involved in the retrieval process and the degree of influence that the grammar has on determining what serves as a cue. As discussed in Chapters 1 and 4, there are two primary accounts regarding the role of grammar in retrieval: unconstrained cues accounts and structurally driven retrieval (e.g. Badecker and Straub, 2002 for unconstrained cues and Chow et al., 2014 for structurally driven retrieval). The primary difference between these two approaches is that the latter limits the set of potential antecedents to those in syntactically licit positions, whereas the former assumes that syntactic constraints act in conjunction with a variety of other factors, whereby allowing some syntactically illicit antecedents to be activated during the retrieval process. Another important distinction between these accounts is that the structure-based accounts primarily focus on tree geometry as a restrictor of antecedents. For the sake of this discussion, I am going to consider any syntactically driven information and biases as of structural involvement in retrieval (e.g., agreement, Locality Bias). This approach varies slightly from the structurally driven cue accounts because it does not focus on tree geometry. With that being said, Chapters 3 and 4 provide clear support for the involvement of syntax in the retrieval process, but these studies were not designed to arbitrate between the two accounts of cues.

Support for structural involvement in retrieval can be found in (1) in the marginal interaction between Number Match and Clause Type in RC Experiment 1, as well as (2) the effects of Locality observed in all of the sluicing experiments. If structural information were not considered during retrieval, there should have been no effect of structural position or bias to change the strength of interference effects. The RCs make a particularly strong case for the involvement of syntactic features because the effect of number match only arises in cases where the matrix and RC subjects match on the syntactic feature of [+Nominative]. This interaction is highly indicative of some sort of structural cue weighting procedure, as discussed in Van Dyke and McElree (2011) and Parker (2019). In the case of sluices, the Locality Bias was so strong that subjects were willing to ignore grammatical information and entertain incorrect targets for retrieval, as shown by the selection of infelicitous distractors in the comprehension question responses. It is still unknown if the Locality Bias in sluicing is purely a structural effect, but regardless, the findings corroborate those found in Harris (2015, 2019) and can be loosely interpreted as a grammatical constraint on cues.

I would also like to revisit the discussion about how antecedents are restricted by addressing the concept of feature *relevance*. To the best of my knowledge, there are no direct predictions about how the retrieval mechanism treats two highly similar morphosyntactic features when they are grammatically required but not necessarily relevant to the dependency in question. More specifically, both number and gender features undergo syntactic agreement procedures in Portuguese, but only number was involved in the retrieval process relevant to most of the grammatical dependencies tested in this thesis.

While Chapters 3 and 4 demonstrate that structural preferences and biases are involved in the retrieval process (e.g., Case in ORCs and the Locality Bias in sluices), they also demonstrate

that *overt* structural information does not always behave in the same manner. In all of the experiments, number was overtly expressed at the retrieval site, making this a relevant feature for retrieval. On the other hand, gender was only overtly marked at the retrieval site in one study (Sluicing Experiment 2). In all other experiments, gender was irrelevant to the grammatical dependency being results. Several clear differences were observed between relevant and irrelevant features, which can potentially be used to further refine our understanding of the linguistic information that acts as a retrieval cue and limits antecedent sets. Retrieval is more strongly influenced by *relevant* features that are expressed at the retrieval site, whereas irrelevant features produce much weaker interference effects that may not derive from the retrieval mechanism at all. This line of argumentation aligns well with the idea that structural information constrains cues, as well as propels the theory of cues forward because it adds yet another way to predict where interference will arise in any dependency. Simply put, if the retrieval cue does not directly project the need for a specific feature, it will not be considered by the retrieval mechanism.

Overall, the findings of these experiments demonstrate that structural preferences influence similarity-based interference effects and that overlapping morphosyntactic information is strong enough to generate processing delays in certain contexts. A further finding was that the results in Chapters 3 and 4 imply that the retrieval mechanism is more sensitive to relevant than irrelevant features, which is a standard assumption in current retrieval theory. Even though the connections between the experiments are not always crystal clear, I hope that this dissertation provides some empirical data that can be used in future discussions about the weighting of structural and morphosyntactic features by the retrieval mechanism.

5.5 Conclusions

As this chapter has demonstrated, the results of the sluices and RCs differed significantly, which made it challenging to provide clear support for any of the four hypotheses presented. However, the findings are consistent with for a combination of two hypotheses: Relevant Feature Retrieval and Encoding Interference. In both the RCs and the sluices, the experiments in which only one relevant feature was manipulated aligns closely with the Relevant Feature Retrieval Hypothesis. These claims become slightly complicated by the combined experiments that crossed Gender and Number, but following the discussions above, due to the lack of number effects in the RCs and the interactions between gender and syntactic information in both the RCs and the sluices. The lack of number interference in RCs can likely be accounted for by experiment design effects and the optionality of overt number expression in Portuguese subject-verb dependencies. Moreover, the interactions that were observed between gender and number, although predicted by several hypotheses, can concretely be explained by the involvement of both encoding and retrieval interference in sentence processing. The encoding hypothesis gained a significant amount of support by the presence of gender match penalties that were before the retrieval site and independent of any clause type or Locality manipulation. The fact that gender match caused reading disruptions and decreased comprehension accuracy in cases where no interference was expected (e.g., SRCs and Local Sluices) aligns more with an encoding approach than with the retrieval hypotheses.

None of these experiments can definitively differentiate between a retrieval mechanism that considers irrelevant features and one that only focuses on relevant features but experiences effects from encoding interference. Although the patterns were not identical across structures, the interference caused by irrelevant features in both sets of experiments requires further research to

determine if they were generated by encoding or retrieval interference. It is not surprising that encoding interference is relatively elusive in many studies, especially considering that the vast amount of psycholinguistic studies that examine English, which does not allow for manipulations between relevant and irrelevant features. Although several studies have failed to identify encoding interference (Jäger et al., 2015; Van Dyke and McElree, 2011), the findings in this study closely align with those of Villata et al. (2018) and can be accounted for by processes such as feature overwriting decreasing the retrievability of targets. This is not to say that the interference generated by irrelevant features could not be accounted for by a more generalized retrieval mechanism, but it would be challenging to explain why irrelevant features cause earlier and behaviorally distinct effects when compared with relevant features. While adopting an encoding + retrieval approach bypasses the need to explain these distinct patterns within the same mechanism, it is an inherently more complex model and much research is needed to disentangle these two approaches.

I would also like to point out that the studies that have failed to find interference were on reflexives and anaphors or semantic information. Anaphors, in particular, have been shown to experience non-standard retrieval interference effects for relevant retrieval features, which suggests that their behavior is not representative of standard cue-based retrieval theory (Dillon et al., 2013). Due to their irregular behavior in retrieval, it is not surprising that they do not demonstrate encoding interference effects. Moreover, the Van Dyke and McElree (2011) study used a completely separate paradigm (memory load task) and were not able to conclusively say that the manipulation did not cause interference effects. The alignment, then, between my studies and those of Villata et al. (2018), who also looked at Romance RCs, should be considered strong support that encoding effects surface, at the very least, in Romance subject-verb dependencies.

In addition to opening the door for future studies on encoding interference, this thesis also helps to further refine what we know about the cue-based retrieval mechanism. By testing the behavior of the same morphological features two separate constructions, it becomes clear that relevant features generate more similarity-based interference effects is than irrelevant features. This helps with future experiment design, as researchers can use this information to make sure that the feature being tested is expected to generate retrieval interference effects in the dependency being tested.

Finally, I provided a brief discussion about the nature of cue-weighting. It appears that relevant features are likely subjects for cue-weighting in a way that is not observed for irrelevant features. If my claims regarding the lower weight for number in Portuguese are corroborated in further work, this assumption will have strong support. Although a significant amount of future research is needed to support these claims, I present these assumptions to hopefully contribute to future discoveries regarding memory processes and sentence processing.

5.6 Future Directions

In the immediate future, I propose several studies than disentangle the results found in this dissertation. First and foremost, I would like to test how speakers of Portuguese that use standard number agreement behave in all of the studies presented in this dissertation. If these studies were to show that number produces stronger interference effects, I would be able to concretely say that the unexptected results of the RC experiments were driven by dialectal cue-weighting. This would also contribute significantly to the existing models of cue-weighting, cross-linguistically, and potentially explain why some features behave differently in different languages.

In addition, all of these studies should be conducted using a more sensitive measure such as eye-tracking. The SPR method employed in these studies was sensitive enough to be able to

capture the Locality Bias and RC asymmetry, which indicate that it was functional; however, some of the encoding and retrieval effects might have been too weak to be captured using this paradigm. Moreover, by not allowing subjects to look back in the sensitive, there may be a completely distinct processing profile of relevant and irrelevant features that was not identifiable in these studies. It would be interesting to be able to comment more directly on the time course of retrieval and potential encoding effects, as this could help to separate my proposed retrieval hypotheses from the encoding + retrieval approach that I have supported in these conclusions.

Finally, there is great value in pursuing psycholinguistic research using non-English languages and hope to see this type of research continue in the future. Regardless of the topic being studied, exploring a variety of languages will help our field move forward and make more generalizable claims. Non-English languages are particularly insightful when looking at morphosyntactic information, feature relevancy, and retrieval/encoding effects, since English contains a relatively impoverished morphological system. As the field progresses, it is crucial that psycholinguists continue to identify the universal components of human cognition that are involved in sentence processing, and without cross-linguistic comparisons, claims of universality become more challenging to process.

Bibliography

Aldridge, E. (2017). Internally and externally headed relative clauses in Tagalog. *Glossa: a journal of general linguistics*, *2*(1).

AnderBois, S. (2010, August). Sluicing as anaphora to issues. In *Semantics and linguistic theory* (Vol. 20, pp. 451-470).

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive psychology*, *6*(4), 451-474.

Anderson, J. R. (1983a). A spreading activation theory of memory. Journal of Verbal Learning and Verbal Behavior, 22(3), 26-295.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. Psychological review, 111(4), 1036.

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. Cognitive science, 29(3), 313-341.

Andrews, A. D. (2007). Relative clauses. Language Typology and Syntactic Description, 2, 206-236.

Arregi, K. (1998). Spanish el que relative clauses and the Doubly Filled COMP Filter. *Ms., MIT*.

Azevedo, M. M. (1984). Loss of agreement in Caipira Portuguese. Hispania, 403-409.

Baddeley, A. D. (1986). Working memory Oxford. England: Oxford Uni.

Baddeley, A. (1996). Exploring the central executive. The Quarterly Journal of Experimental Psychology Section A, 49(1), 5-28.

Baddeley, A. D. (2002). Is working memory still working?. European psychologist, 7(2), 85.

Baddeley, A. (2003). Working memory and language: An overview. Journal of Communication Disorders, 36(3), 189-208.

Baddeley, A., & Hitch, G. (1972). Working memory. GM BOWER. The psychology of learning and motivation, 8.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In Psychology of learning and motivation (Vol. 8, pp. 47-89). Elsevier.

Baddeley, A. D., & Della Sala, S. (1996). Working memory and executive control. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 351(1346), 1397-1404.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(4), 748.

Barker, C. (2013). Scopability and sluicing. Linguistics and Philosophy, 36(3), 187-223.

Barros, M. (2014). Sluicing and identity in ellipsis (Doctoral dissertation, Rutgers University-Graduate School-New Brunswick).

Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31. *CRAN. R-project. ore/packaee= lme4*.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, *12*(1), 2.

Bhatt, R. (2002). The raising analysis of relative clauses: Evidence from adjectival modification. *Natural language semantics*, *10*(1), 43-90.

Belletti, A., Friedmann, N., Brunato, D., & Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on childrenn's comprehension of relative clauses in Hebrew and Italian. Lingua, 122(10), 1053-1069.

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

Betancort, M., Carreiras, M., & Sturt, P. (2009). Short article: The processing of subject and object relative clauses in Spanish: An eye-tracking study. Quarterly Journal of Experimental Psychology, 62(10), 1915-1929.

Bever, T. G. (1970). The cognitive basis for linguistic structures. Cognition and the development of language, 279(362), 1-61.

Bever, T. G. (1974). The ascent of the specious, or there's lot we don't know about mirrors. Explaining Linguistic Phenomena, 173-200.

Bhatt, R. (2002). The Raising Analysis of Relative Clauses: Evidence from Adjectival Modification, 48.

Brame, M. K. (1968). On the Nature of Relative Clauses. *Unpublished paper, MIT*.

Brown, J. (1958). Some tests of the decay theory of immediate memory. Quarterly Journal of Experimental Psychology, 10(1), 12-21.

Carlson, K., Dickey, M. W., Frazier, L., and Clifton, C. Jr. (2009). Information structure expectations in sentence comprehension. Q. J. Exp. Psychol. 62, 114-139.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology Section A*, *49*(3), 639-663.

Chen, Z., Jäger, L., & Vasishth, S. (2012). How structure-sensitive is the parser? Evidence from Mandarin Chinese. Empirical approaches to linguistic theory: Studies of meaning and structure, 43-62.

Chomsky, N. (1957). Syntactic structure. Mouton.

Chomsky, N. (1959). Verbal behavior by BF Skinner. Bobbs-Merrill.

Chomsky, N. (1965). Aspects ofthe theory of syntax. *Cambridge, MA: MITPress*.

Chomsky, N., & Lasnik, H. (1977). Filters and control. Linguistic Inquiry, 8(3), 425-504.

Chow, W. Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. Frontiers in psychology, 5.

Corbett, G. G. (2005). Systems of gender assignment (chapter and map).

Chung, C., & Kim, J.-B. (2003). Differences between externally and internally headed relative clause constructions. In On-line Proceedings of HPSG 2002 (pp. 3-25). Citeseer.

Chung, S., Ladusaw, W. A., & McCloskey, J. (1995). Sluicing and logical form. Natural language semantics, 3(3), 239-282.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. Psychonomic bulletin & review, 3(1), 37-60.

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In Linguistic structure in language processing (pp. 273-317). Springer.

Cole, P. (1987). The structure of internally headed relative clauses. Natural Language and Linguistic Theory, 5(2), 277-302.

Conway, A. R., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. Journal of Experimental Psychology: General, 123(4), 354.

Corbett, G. G. (2005). Systems of gender assignment (chapter and map).

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and brain sciences, 24(1), 87-114.

Cowan, N. (2016). Working memory capacity: Classic edition. Psychology press.

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

Dillon, B. W. (2011). *Structured access in sentence comprehension* (Doctoral dissertation).

Dillon, B. (2014). Syntactic memory in the comprehension of reflexive dependencies: an overview. Language and Linguistics Compass, 8(5), 171-187.

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., 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.

Dixon, W. J. (1960). Simplified estimation from censored normal samples. The Annals of Mathematical Statistics, 385-391.

Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. Studies in Second Language Acquisition, 13(4), 431-469.

Eckman, F. R., Bell, L., & Nelson, D. (1988). On the Generalization of Relative Clause Instruction in the Acquisition of English as a Second Language1. Applied Linguistics, 9(1), 1-20.

Engle, R. W. (1996). Working memory and retrieval: An inhibition-resource approach. Working Memory and Human Cognition, 89119.

Ericsson, K. A., & Pennington, N. (1993). The structure of memory performance in experts: Implications for memory in everyday life. In Advances in psychology (Vol. 100, pp. 241-272). Elsevier.

Estivalet, G. L., & Meunier, F. Corpus psicolinguístico Léxico do Português Brasileiro. Revista SOLETRAS, (33), 212-229.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. Journal of Memory and Language, 54(4), 541-553

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current directions in psychological science*, *11*(1), 11-15.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*(3), 427-473.

Foraker, S., & McElree, B. (2011a). Comprehension of Linguistic Dependencies: Speed-Accuracy Tradeoff Evidence for Direct-Access Retrieval From Memory. Language and Linguistics Compass, 5(11), 764-783.

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. Journal of Memory and Language, 22(2), 203.

Franck, J., Colonna, S., & Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. Frontiers in psychology, 6, 349.

Franck, J., & Wagers, M. (2020). Hierarchical structure and memory mechanisms in agreement attraction. *Plos one*, *15*(5), e0232163.

Frank, R. E. (1992). Syntactic locality and tree adjoining grammar: grammatical, acquisition and processing perspectives.

Frazier, L. (1978). On comprehending sentences: Syntactic parsing strategies. *Doctoral dissertation, University of Connecticut*.

Frazier, L. (1987). Syntactic processing: evidence from Dutch. Natural Language & Linguistic Theory, 5(4), 519-559.

Frazier, L., & d'Arcais, G. B. F. (1989). Filler driven parsing: A study of gap filling in Dutch. Journal of memory and language, 28(3), 331-344.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. Cognition, 6(4), 291-325.

Frazier, L., & Clifton Jr, C. (1989). Successive cyclicity in the grammar and the parser. Language and Cognitive Processes, 4(2), 93,-126.

Frazier L., Clifton C., Jr. (1998). Comprehension of sluiced sentences. Lang. Cogn. Process. 13, 499,-520.

Frazier, L., & Clifton, C. (1998). Sentence reanalysis, and visibility. In Reanalysis in sentence processing (pp. 143-176). Springer, Dordrecht.

Frazier, L., Clifton, C., & Randall, J. (1983). Filling gaps: Decision principles and structure in sentence comprehension. Cognition, 13(2), 187-222.

Garrod, S., & Sanford, A. (1977). Interpreting Anaphoric Relations: The Integration of Semantic. Journal of Verbal Learning and Verbal Behavior, 16, 77-90.

Garrod, S. C., & Sanford, A. J. (1982). The mental representation of discourse in a focused memory system: Implications for the interpretation of anaphoric noun phrases. Journal of semantics, 1(1), 21-41.

Garrod, S. C., & Sanford, A. J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding.

Gibson, E. A. F. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown (PhD Thesis). Carnegie Mellon University Pittsburgh, PA.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. Cognition, 68(1), 1,-76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, *2000*, 95-126.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. Language and Cognitive Processes, 14(3), 225,-248.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(6), 1411

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. Psychological science, 13(5), 425-430.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. Journal of Memory and Language, 51(1), 97,-114.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-Based Interference During Language Comprehension: Evidence from Eye Tracking During Reading. Journal of Experimental Psychology. Learning, Memory, and Cognition, 32(6), 1304,-1321.

Gouvea, A. C. de S. L. (2003). Processing syntactic complexity: Cross-linguistic differences and ERP evidence (PhD Thesis). University of Maryland, College Park.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive science*, *29*(2), 261-290.

Hakuta, K. (1981). Grammatical description versus configurational arrangement in language acquisition: The case of relative clauses in Japanese. Cognition, 9(3), 197,-236. https://doi.org/10.1016/0010-0277(81)90001-9

Hale, J. (2001). A Probabilistic Early Parser as a Psycholinguistic Model. In Second Meeting of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics.

Halliday, M. A. K., & Hasan, R. (1976). Cohesion in English. London: Longman.

Haviland, S.E. & H.H. Clark (1974). What's new? Acquiring new information as a process in comprehension. Journal of Verbal Learning and Behavior. 13:512-521.

Harris, Jesse A. & Katy Carlson (2016). Keep it local (and final): Remnant preferences for let alone ellipsis. Quarterly Journal of Experimental Psychology, 69: 1278-1301.

Harris, J. A. (2015). Structure Modulates Similarity-Based Interference in Sluicing: An Eye Tracking study. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.01839

Harris, J. A., & Carlson, K. (2018). Information structure preferences in focus-sensitive ellipsis: How defaults persist. *Language and speech*, *61*(3), 480-512.

Harris, Jesse A. (2019). Alternatives on demand and locality: Resolving discourse-linked wh-phrases in sluiced structures. In: Clifton, C., Fodor, J.D., & Carlson, K. Grammatical Approaches to Language Processing - Essays in Honor of Lyn Frazier. Springer Studies in Theoretical Linguistics.

Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 417-430.

Hulsey, S., & Sauerland, U. (2006). Sorting out Relative Clauses. Natural Language Semantics, 14(2), 111,-137. https://doi.org/10.1007/s11050-005-3799-3

Hunter, T. (in progress). Formal Methods in Experimental Syntax. In The Oxford Handbook of Experimental Syntax (J. Sprouse (ed.), p. 74). Retrieved from https://linguistics.ucla.edu/people/hunter/parsing/expt-syntax-handbook.pdf

Ishizuka, T., Nakatani, K., & Gibson, E. (2006). Processing Japanese relative clauses in context. In 19th annual CUNY conference on human sentence processing.

Jäger, L. A., Benz, L., Roeser, J., Dillon, B. W., & Vasishth, S. (2015). Teasing apart retrieval and encoding interference in the processing of anaphors. Frontiers in psychology, 6.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. Journal of Memory and Language, 94, 316,-339.

Kato, M., & Nunes, J. (2009). A uniform raising analysis for standard and nonstandard relative clauses in Brazilian Portuguese. Minimalist Essays on Brazilian Portuguese Syntax, 93,-120.

Kenedy, E. (2002). Aspectos estruturais da relativação em português: uma análise baseada no modelo raising (PhD Thesis). Dissertação (Mestrado),Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro.

Kenedy, E. (2014). Estruturas sintáticas de orações relativas. *BISPO, EB & DE OLIVEIRA, MR*.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. Journal of verbal learning and verbal behavior, 1(3), 153-161.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. Cognition, 2(1), 15,-47.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. Journal of Cognitive Neuroscience, 7(3), 376,-395.

Kush, D., Johns, C. L., & Van Dyke, J. A. (2015). Identifying the role of phonology in sentence-level reading. *Journal of memory and language*, *79*, 18-29.

Laurinavichyute, A., Jäger, L. A., Akinina, Y., Roß, J., & Dragoy, O. (2017). Retrieval and encoding interference: Cross-linguistic evidence from anaphor processing. *Frontiers in psychology*, *8*, 965.

Lawn and Harris (in progress). Sluicing and retrieval in Brazilian Portuguese.

Lees, R. B. (1960). The grammar of english nominalizations: Indiana university. research center in anthropology. Folklore and Linguistics, Bloomington, Indiana, 10.

Lewandowsky, S., Oberauer, K., & Brown, G. D. (2009). No temporal decay in verbal short-term memory. *Trends in cognitive sciences*, *13*(3), 120-126.

Levy, R. (2005). Probabilistic models of word order and syntactic discontinuity. stanford university.

Levy, R. (2008). Expectation-based syntactic comprehension. Cognition, 106(3), 1126,-1177.

Levy, R., & Gibson, E. (2013). Surprisal, the PDC, and the primary locus of processing difficulty in relative clauses. Frontiers in Psychology, 4, 229.

Lewis, R. L. (1993). An architecturally-based theory of human sentence comprehension. In Proceedings of the fifteenth annual conference of the cognitive science society (pp. 108,-113).

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

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, *29*(3), 375-419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. Trends in Cognitive Sciences, 10(10), 447,-454.

MacWhinney, B., & Pleh C. (1988). The processing of restrictive relative clauses in Hungarian. Cognition, 29(2), 95,-141.

MacWhinney, B. (1997). Second language acquisition and the competition model. Tutorials in bilingualism: Psycholinguistic perspectives, 113-142.

Makovski, T., & Jiang, Y. V. (2009). The role of visual working memory in attentive tracking of unique objects. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1687.

Marantz, A. (1984). On the Nature of Grammatical Relations, vol. 10 of Linguistic Inquiry Monographs.

Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. Frontiers in psychology, 7, 120.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. Journal of Memory and Language, 58(3), 879-906.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: evidence from verb-phrase ellipsis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(5), 1231.

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: evidence from sluicing. Journal of memory and language, 64(4), 327-343.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. Journal of Experimental Psychology: General, 118(4), 346,-373.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. Journal of Experimental Psychology: General, 122(3), 291.

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(1), 134.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. Journal of memory and language, 48(1), 67-91.

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. Memory & Cognition, 24(2), 173-187.

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. (2000). Sentence comprehension is mediated by content-addressable memory structures. Journal of psycholinguistic research, 29(2), 111-123.

McElree, B. (2001). Working memory and focal attention. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(3), 817,-835.

McElree, B. (2006). Accessing recent events. Psychology of learning and motivation, 46, 155-200.

Merchant, J. (2001). The syntax of silence: Sluicing, islands, and the theory of ellipsis. Oxford University Press on Demand.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users.

Miyamoto, E. T., & Nakamura, M. (2003). Subject/object asymmetries in the processing of relative clauses in Japanese. In Proceedings of the 22nd West Coast conference on formal linguistics (Vol. 22, pp. 342,-355). Cascadilla Press Somerville, MA.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251-269.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. Journal of psycholinguistic research, 18(1), 5-19.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. Journal of memory and language, 55(4), 601-626.

Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. Journal of Memory and Language, 58(3), 730-745.

Oberauer, K., & Lewandowsky, S. (2013). Evidence against decay in verbal working memory. *Journal of Experimental Psychology: General*, *142*(2), 380.

Oberauer, K., & Hein, L. (2012). Attention to Information in Working Memory. Current Directions in Psychological Science, 21(3), 164,-169.

Parker, D. (2019). Cue combinatorics in memory retrieval for anaphora. Cognitive science, 43(3), e12715.

Parker, D., Lago, S., & Phillips, C. (2015). Interference in the processing of adjunct control. Frontiers in psychology, 6, 1346.

Parker, D., & Phillips, C. (2016). Reflexive attraction is selective: Evidence from eye tracking. *Manuscript submitted for publication*.

Parker, D., MacMillan, G., & Phillips, C. (2013). Illusory NPI licensing: now you see it, now you don't. In Poster Presented at the 26th Annual CUNY Human Sentence Processing Conference. Columbia, SC: University of South Carolina.

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2017). The cue-based retrieval theory of sentence comprehension: New findings and new challenges. Language processing and disorders, 121-144.

Partee, B. (1975). Montague grammar and transformational grammar. *Linguistic inquiry*, 203-300.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. Journal of Experimental Psychology, 58(3), 193.

Plann, S. (1980). Relative clauses in Spanish without overt antecedents and related constructions. University of California Publications in Linguistics Berkeley, Cal., 93, 1,-196.

Plann, S. J. (1975). Relative Clauses in Spanish.

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

R Development Core Team (2014). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Ross, J. R. (1969). Guess Who. In et al. Binnick, RI, editor. In Fifth Regional Meeting of the Chicago Linguistics Club.

Santorini, B., & Kroch, A. (2007). The syntax of natural language: An online introduction using the Trees program. Retrieved June, 24, 2011.

Sasson, R. Y. (1971). Interfering images at sentence retrieval. Journal of Experimental Psychology, 89(1), 56.

Sauerland, U. (2003). Unpronounced heads in relative clauses. The Interfaces: Deriving and Interpreting Omitted Structures, 61, 205,-226.

Sauerland, U. (2004). The interpretation of traces. Natural Language Semantics, 12(1), 63,-127.

Sakamoto, Y., Ishiguro, M., and Kitagawa G. (1986). Akaike Information Criterion Statistics. D. Reidel Publishing Company.

Schachter, P. (1973). Focus and relativization. *Language*, 19-46.

Schlesinger, I. M. (1975). Why a sentence in which a sentence in which a sentence is embedded is embedded is difficult. Linguistics, 13(153), 53,-66.

Schlueter, Z., Parker, D., & Lau, E. F. (2019). Error-driven Retrieval in Agreement Attraction rarely leads to Misinterpretation. *Frontiers in psychology*, *10*, 1002.

Schneider, W., & Detweiler, M. (1988). A connectionist/control architecture for working memory. In Psychology of learning and motivation (Vol. 21, pp. 53,-119). Elsevier.

Schriefers, H., Friederici, A. D., & Kuhn, K. (1995). The processing of locally ambiguous relative clauses in German. Journal of Memory and Language; New York, 34(4), 499,-520.

Schlueter, Z., Parker, D., & Lau, E. F. (2019). Error-driven Retrieval in Agreement Attraction rarely leads to Misinterpretation. *Frontiers in psychology*, *10*, 1002.

Shallice, T., & Vallar, G. (1990). The impairment of auditory-verbal short-term storage. Neuropsychological Impairments of Short-Term Memory, 11,-53.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. Journal of Verbal Learning and Verbal Behavior, 13(3), 272,-281.

Sportiche, D., Koopman, H., & Stabler, E. (2013). An introduction to syntactic analysis and theory. John Wiley & Sons.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. Cognition, 116(1), 71,-86.

Stavrakaki, S. (2001). Comprehension of Reversible Relative Clauses in Specifically Language Impaired and Normally Developing Greek Children. Brain and Language, 77(3), 419,-431.

Sternberg, S. (1966). High-speed scanning in human memory. Science, 153(3736), 652-654.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist, 57(4), 421,-457.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. The Quarterly journal of experimental psychology, 27(1), 1-32.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. Language and cognitive processes, 1(3), 227-245.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. Journal of Memory and Language, 48(3), 542-562.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. Journal of Memory and Language, 47(1), 69,-90.

Tukey, J. W. (1962). The future of data analysis. The annals of mathematical statistics, 33(1), 1-67.

Ueno, M., & Garnsey, S. M. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. Language and Cognitive Processes, 23(5), 646,-688.

Van Dyke, J. (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. https://doi.org/10.1016/S0749-596X(03)00081-0

Van Dyke, J. A. (2011). Cue-dependent interference in comprehension. Journal of Memory and Language, 65(3), 247,-263.

Van Dyke, J. A., & Johns, C. L. (2012). Memory Interference as a Determinant of Language Comprehension. Language and Linguistics Compass, 6(4), 193,-211.

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. Cognition, 131(3), 373-403.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. Journal of Memory and Language, 49(3), 285-316.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. Journal of Memory and Language, 55(2), 157,-166.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. Journal of memory and language, 65(3), 247-263.

Vasishth, S., Jäger, L. A., & Nicenboim, B. (2017). Feature overwriting as a finite mixture process: Evidence from comprehension data. *arXiv preprint arXiv:1703.04081*.

Villata, S., & Franck, J. (2016). Attraction and similarity-based interference in object gender agreement. In *CUNY (Conference on Human Sentence Processing)*.

Villata, S., & Franck, J. (2019). Similarity-based interference in agreement comprehension and production: Evidence from object agreement. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. Frontiers in psychology, 9, 2.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. Journal of Memory and Language, 61(2), 206,-237.

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. Linguistic theory and psychological reality, ed. by Morris Halle, Joan Bresnan and George Miller, 119,-61. Cambridge, MA: MIT Press.

Warren, T., & Gibson, E. (1999). The effects of discourse status on intuitive complexity: Implications for quantifying distance in a locality-based theory of linguistic complexity. In Poster presented at the Twelfth CUNY Sentence Processing Conference, New York.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 442.

Watkins, M. J., & Watkins, O. C. (1976). Cue-overload theory and the method of interpolated attributes. Bulletin of the Psychonomic Society, 7(3), 289,-291.

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. Psychophysiology, 36(5), 559,-570.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. Brain and Language, 108(1), 40-55.

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. Language and Cognitive Processes, 28(3), 272-302.