Federal University of Rio de Janeiro
Faculty of Letters
Linguistics Graduate Program

The influence of gender cues in pronominal antecedent retrieval in Brazilian Portuguese

By

Michele Calil dos Santos Alves

2018

The influence of gender cues in pronominal antecedent retrieval in Brazilian Portuguese

By

Michele Calil dos Santos Alves

Doctoral dissertation presented to the Linguistics Graduate Program at the Federal University of Rio de Janeiro in partial fulfillment of the requirements for the degree of Doctor in Linguistics.

Adviser: Marcus Antonio Rezende Maia

Rio de Janeiro,

March 2018

Alves, Michele Calil dos Santos
The influence of gender cues in pronominal antecedent retrieval in Brazilian Portuguese / Michele Calil dos Santos Alves. – Rio de Janeiro: UFRJ/Faculty of Letters 2018.

viii, 300 f.:I1; 31 cm.
Adviser: Marcus Antonio Rezende Maia
Doctoral dissertation (PhD) – UFRJ/ Faculty of Letters 2018.

References: f. 288-296
1. Pronominal antecedent retrieval processing; Memory retrieval; Gender cues; Brazilian Portuguese; Experimental Psycholinguistics
I. Maia, Marcus Antonio Rezende (Adviser). II. Federal University of Rio de Janeiro. Faculty of Letters. Linguistics Graduate Program. III. The influence of gender cues in pronominal antecedent retrieval in Brazilian Portuguese.

ALVES, Michele Calil dos Santos. The influence of gender cues in pronominal antecedent retrieval in Brazilian Portuguese / Federal University of Rio de Janeiro, Faculty of Letters, 2018. Doctoral Dissertation in Linguistics.

DOCTORAL DISSERTATION DEFENSE COMMITEE

_____

President (Advisor): Marcus Antonio Rezende Maia
Federal University of Rio de Janeiro

_____

Internal Member: Aniela Improta França
Federal University of Rio de Janeiro

_____

Internal Member: Daniela Cid de Garcia
Federal University of Rio de Janeiro

_____

External Member: Letícia Maria Sicuro Corrêa
Catholic University of Rio de Janeiro

_____

External Member: Elisângela N. Teixeira
Federal University of Ceará

ACKNOWLEGEMENTS

I am very thankful to God for all the blessings He has generously given me throughout this journey.

I am very thankful to my family for all the support they have given me to study since my childhood.

I am very thankful to my husband for loving and believing in me. You are the best partner someone could ever have.

I am very thankful to my advisor Marcus Maia for trusting and guiding me.

I am very thankful to my colleagues for all the inspiration they gave me: Katharine, Priscila, Cristiane, Ana, Thiago, Nathacia, Sara, Lorrane, Gitanna, and Rafaela.

I am very thankful to Colin Phillips for accepting me in my doctoral internship at the University of Maryland, where I learned so much.

I am very thankful to Ellen Lau for the wonderful classes and the suggestions she gave to my projects.

I am very thankful to my UMD colleagues for receiving me so well.

I am very thankful to my Fulbright colleagues for making me a better person and for the fantastic time we spent together.

I am very thankful to Aniela, Daniela, Letícia, and Elisângela for participating in my PhD not only as examiners, but truly helping me to improve my research.

I am very thankful to CNPq, CAPES, and Fulbright for the financial aid and the opportunity of doing my research.

And last, but not least, I am very thankful to the many volunteers who have participated in my experiments.

ABSTRACT

The influence of gender cues in pronominal antecedent retrieval in Brazilian Portuguese

MICHELE CALIL DOS SANTOS ALVES

Adviser: Marcus Antonio Rezende Maia

Abstract of the doctoral dissertation submitted to the Program of Graduation in Linguistics at the Federal University of Rio de Janeiro – UFRJ, in order to receive the title of Doctor in Linguistics.

Coreference is a syntactic dependency in which pronouns are bound to previous referents in discourse. One of the keys factors influencing coreference processing is memory, since information that has already been interpreted and stored must be integrated with new material in real time. Granted that antecedents of anaphors must be retrieved from memory in coreference, the aim of this research is to provide more information on how pronominal antecedents are retrieved, and more precisely to clarify the role of gender cues in pronominal antecedent retrieval when gender morphology is overt. Since Brazilian Portuguese is a language with visible morphology, speakers of this language are used to rely on agreement cues to process language. Thus the first hypothesis of the present research is gender morphological cues play a great role in pronominal antecedent retrieval in Brazilian Portuguese. This way, structurally unacceptable antecedent candidates that agree in gender with the pronouns would be considered as potential candidates, despite the fact they violate pronominal binding constraints. The second hypothesis of the present research is memory can be influenced by language so that different gender features might assume different weights (van Dyke & McElree, 2011) in memory. This way, memory is affected by the prominence of gender features. The results of four eye-tracking

experiments conducted with native speakers of Brazilian Portuguese demonstrated that both binding structural constraints and gender morphological cues are equally important in antecedent retrieval in memory throughout processing. In addition, the results indicated that types of gender that are lexically determined seem to be more prominent in memory than other types of gender such as the compositional/derivational gender and the syntactic gender.

Table of Contents

List of Tables

List of Figures

1. Introduction

In order to process language in real time, previously interpreted information must be kept at least momentarily in memory so that integration with novel upcoming material can take place rapidly (Lewis et al., 2006). This way, memory can be considered one of the key factors in processing long distance dependencies such as coreference, in which pronouns are bound to antecedents that occupy linearly distant positions in the discourse.

Among other cues, coreference can be influenced by binding structural constraints, agreement relations between antecedents and pronouns, and salience of the discourse entities involved in the context. Previous research that has investigated how those three factors play a role in binding processing is very contradictory. On the one hand, it has been claimed that structurally unacceptable candidates cannot initially influence binding processing even in cases in which they are salient discourse entities and agree with the anaphoric expressions [Nicol & Swinney, 1989; Clifton et al., 1997; Sturt, 2003; Leitão et al., 2008; Xiang et al., 2009; Oliveira et al., 2012; Dillon et al., 2013; Chow et al., 2014]. On the other hand, other research has shown that structural constraints can be fallible as apparently structurally unacceptable candidates can be initially considered as potential antecedents if they are salient entities that feature-match the anaphors [Badecker & Straub, 2002; Kennison, 2003; Parker, 2014; Patil et al., 2016].

One possible explanation for these contradictory results in relation to the role of agreement in binding processing in the literature may rely on the fact that those studies may have not taken into account intrinsic differences that exist among morphological features. Languages with limited overt morphology like English might

not be the most appropriate to study gender agreement. By comparing overt agreement marking in English and in Portuguese, one notices that unlike the former, the latter has redundant gender agreement marking in most determiners, nouns, and adjectives, for example. In these terms, the present study tried to control for the different types of features that may exist under the category of gender. This way, this dissertation aims at verifying whether different types of gender conveyed by pronominal antecedent candidates would influence the way they would be retrieved from memory.

Thus this dissertation focuses on Portuguese whose redundant overt morphology of agreement cues may be more fruitful. Moreover, it seems that the use of morphological cues in memory retrieval may also vary depending on the particular binding dependency. Agreement features may be more helpful in pronominal antecedent retrieval due to the looseness of its constraints. In other words, pronominal binding constraints (Principle B of Binding Theory) only posits antecedents must not be local, which is not a quite restrictive constraint. This way, morphological cues would be very helpful in pronominal antecedent retrieval.

This way, the present research will fill a gap in the literature as it will provide not only one more piece of evidence to the puzzle of binding processing, which lacks intensive investigation, but it will provide evidences of the role of gender cues in pronominal binding in a language with redundant visible gender morphology like Brazilian Portuguese. It will also be determined whether speakers of Brazilian Portuguese tend to initially consider structurally unacceptable candidates as potential antecedents despite the fact that they violate binding constraints.

(1) O arquiteto agradeceu o engenheiro que indicou **ele** justamente para um dos cargos mais cobiçados do país.

(The architect$_{[masc]}$ thanked the engineer$_{[masc]}$ who fairly recommended **him** for one of the most desirable jobs in the country.)

For instance, in (1), according to the pronominal binding structural constraints, the pronoun *ele* (him) refers to *arquiteto* (architect$_{[masc]}$); however, there is another antecedent candidate in the sentence that also agrees in gender with the pronoun, *engenheiro* (engineer$_{[masc]}$). The question is whether structurally unacceptable antecedents such as *engenheiro* (engineer$_{[masc]}$) would influence antecedent retrievals in memory. Moreover, another question is whether different genders (masculine or feminine) or even whether different types of gender (compositional/derivational, syntactic, lexical, or stereotypical) conveyed by antecedent candidates would be responsible for any differences in the how antecedents would be retrieved.

In Portuguese, most nouns with semantic gender vary in gender through compositional/derivational processes, for example, *arquiteto* (male architect) versus *arquiteta* (female architect), or *europeu* (male European) versus *européia* (female European). However, there are other different types of gender variation. There are nouns whose gender is syntactically determined such as the epicenes, for example, *vítima* (victim$_{[fem]}$), which is grammatically feminine, but can refer to either a male or a female referent. Moreover, there are nouns with lexically determined gender variation since there are no formal morphological gender cues to indicate gender, as for example, *mulher* (woman) versus *homem* (man). Finally, a third type of nouns is the bigenders, which are gender ambiguous and dependent on context, as for example, *turista* (male or female tourist) or *estudante* (male or female student). Some bigender

nouns are stereotyped biased, for example *recepcionista* (receptionist) is feminine biased while *surfista* (surfist) is masculine biased. This way, taken into account the richness of gender variation in Portuguese, do different types of gender have different weights in memory, that is, different prominence levels in memory?

The aim of this research is to provide more information on how pronominal antecedents are retrieved from memory, and more precisely to clarify the role of gender cues in pronominal antecedent retrieval when gender morphology is overt. Since Brazilian Portuguese is a language with overt morphology, speakers of this language are used to rely on agreement cues to process language. Thus the first hypothesis is gender morphological cues play a great role in pronominal antecedent retrieval in Brazilian Portuguese. This way, structurally unacceptable antecedent candidates that agree in gender with the pronouns would be considered as potential candidates, despite the fact they violate pronominal binding constraints. The second hypothesis is related to the fact different gender features would be encoded/retrieved in memory with different weights (van Dyke & McElree, 2011). Thus memory can be affected by the prominence of gender features.

In order to test the hypotheses, two pre-tests and four eye-tracking experiments were conducted with native speakers of Brazilian Portuguese. The eye-tracking technique is suitable for our purposes as it enables the researcher to examine the temporal course of language processing, including early and late processing measures.

This dissertation will be arranged as follows: chapter 2 will present binding theories and some experimental evidences of binding processing, followed by a brief review of memory models. Then, there will be a discussion on previous research on the interference of structurally unacceptable candidates in binding processing; chapter

3 will address gender in theory and in word processing, followed by a discussion on how different types of gender influence sentence processing; chapter 4 will report and discuss the results of the pre-tests and the experiments conducted in this dissertation; and chapter 5 will summarize the main findings of this dissertation followed by a concluding remarks on the importance of the present research for the field. The last part of this work contains the references and the appendix.

2. Binding in theory and processing

The phenomenon of coreference – where two linguistic expressions refer to the same thing – is a central topic in attempts to understand the meaning and structure of language. This is true both for disciplines that attempt to model knowledge of language and for disciplines that attempt to model language processing
(Gordon & Hendrick, 1998, p.389)

2.1 Outline

In the first section of this chapter, "Binding", the Standard Binding Theory and two Predicate-Based Binding Theories are presented and compared to each other. Experimental data in Brazilian Portuguese that support the Standard Binding Theory will also be presented. Second section, "Pronouns and discourse", will address the importance of discursive factors in pronouns resolution. Theories that explain pronoun resolution with world knowledge inferences, topichood bias, and discursive prominence will be presented, followed by a few experimental data in Brazilian Portuguese. In the third section, "Pronouns, discourse, and gender", the discussion will be turned to gender. Some experimental studies that have dealt with discourse and gender information will be reviewed. In the fourth section, "Pronouns interpretation", the deficiencies of the ordinary binding model used in the literature will be discussed and a more integrative discourse model will be proposed. In the fifth section, "Memory retrieval", the focus of discussion will be turned to the importance of memory in binding processing. Different types of memory models will be reviewed accompanied by some experimental data. In the last section, "Structural and morphological cues in binding processing", the relationship between structural cues of Principle A and B, and morphological cues (especially those involving gender) will be

addressed by a review of the most important and relevant previous works in the literature of interference effects, which is the main object of study of this dissertation.

2.2 Binding

In this section, binding theory will be presented in two flavors: the Standard Binding Theory and the Predicate-Based Binding Theory. The former was proposed by Chomsky (1981, 1986, 1993), while the latter was developed by, among other linguists, Pollard & Sag (1992) and Reinhart & Reuland (1993). A comparison between these two types of theories will be provided as well as a subsection with experimental evidences in favor of the Standard Binding Theory in Brazilian Portuguese.

2.2.1 The Standard Binding Theory

According to several authors, one of the central properties of natural languages is being capable of referring to entities in the world such as people, objects, actions, qualities, etc (cf. Kenedy, 2016). The entities to which the linguistic expressions refer to are called referents. For example, in the sentence “John bought the Linguistics book.”, the person named John is the referent of the term “John”, and the specific book to which the sentence refers to is the referent of the phrase “the Linguistics book”. Interestingly more than one linguistic expression can have the same referent, in this case it is said they are co-referents. For instance, in the sentence, “John said he bought the Linguistics book”, “John” and “he” might be the same

person, and in this case it is said they are co-referents[1]. The relation between a new constituent that refers to another constituent already mentioned in the sentence or in the discourse in said to be "anaphoric". In other words, an anaphoric relation is when a constituent retrieves another constituent previously introduced in discourse.

Linguists have been trying to understand how reference, co-reference, and anaphoric relations are computed by the human mind. One of those linguists is Noam Chomsky, who postulated the binding theory in the 1980s.

(2) Binding: α binds β if α c-commands and is co-indexed with β.

C-command: a node α c-commands β if the node which immediately dominates α also dominates β.

In the tree scheme below, "John" c-commands "he" because IP, which is the first node that dominates "John", also dominates "he". Consequently, "John" is bound to "he" and must be co-indexed. In (3), the index "i" shows that "John" and "he" are co-referents.

(3) $John_i$ said he_i bought the linguistics book.

[1] "He" can also refer a referent not mentioned before; however, this seems awkward, as comprehenders tend to seek for a referent within the given discourse.m

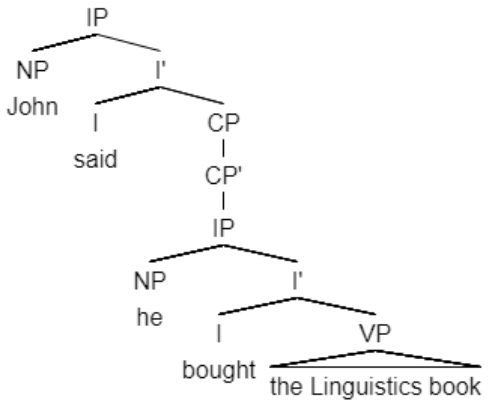

Figure 1: Example of c-command in binding theory

Chomsky (1981 & 1986, reviewed in 1993) postulated three principles in the binding theory: A, B, and C, which are able to explain, respectively, the distributional constraints of: (a) anaphoras, which according to Chomsky, only includes the reflexives ("himself", "herself" etc.) and reciprocals ("each other"); (b) pronouns ("he", "she", etc.); and (c) referential-expressions (which are neither anaphoric nor pronominal, as for example, "John" and "book").

Chomsky claimed that depending on the nature of the NPs involved and the syntactic configurations in which they occur, the anaphoric relations can be possible, necessary, or proscribed.

Referential expressions (R-expressions) do not need antecedents; they need to be free independently of c-command relations. In (4), "John" and "the girl" are R-expressions. Even though, "John" c-commands "the girl" they cannot be bound.

(4) John criticized the girl.

(5) If β is not bound, β is free.

(6) A referential expression must be free (Principle C)

In (7), “John” can be the referent of “him”; on the contrary, in (8), this is not possible.

(7) John$_i$ said Mary criticized him$_i$.

(8) *John$_i$ criticized him$_i$.

According to Chomsky, pronouns and their antecedents cannot be too close from each other. In (7), “John” can be the antecedent of “him” because, unlike (8), there is a clausal border intervening between the pronoun and the antecedent. However, the linear distance is not always enough to allow binding as in (9).

(9) *He$_i$ said Mary criticized John$_i$.

Since it is not a matter of linear distance, pronouns can actually linearly precede their antecedents, like in (10), a construction traditionally known as *cataphora*.

(10) After he$_i$ entered the room, John$_i$ sat down.

Chomsky proposed a pronoun couldn’t take as its antecedent an element within its (c-command) domain. A c-command domain of an element is the minimum constituent that contains this element. In (8), the domain is the whole sentence, and since the antecedent is located within this domain, “John” cannot be the referent of the pronoun “him”. However, in (10), the pronoun domain in within the adverbial

clause, which does not include “John”; therefore, “John” can be the antecedent of “he”.

In (8) and (9), “John” cannot be the referent of “him” and “he” respectively; consequently, they cannot be co-indexed ($i \neq j$).

(11) If the α index is different from the β index, α cannot be the antecedent of β and β cannot be the referent of α.

The example in (8) and (9) would be:

(12) $John_i$ criticized him_j.

(13) He_i said Mary criticized $John_j$.

In (15), one can see that a pronoun can exist within its antecedent domain; however, it cannot be too close to it.

(14) A pronoun must be free in a local domain (Principle B).

The local domain is generally the minimum clause, that is, within the domain of the minimum tense phrase (TP) that contains the pronoun. Despite “John” and “him” are in the same local domain in (15), “John” and “him” can be bound, because “John” does not c-command “him”: the first node that domains “John” is a noun phrase (NP), and this NP does not domain “him”. In other words, pronouns and antecedents can be located in the same clause, but the antecedents cannot c-command the pronouns.

(15) John's$_i$ boss criticized him$_i$

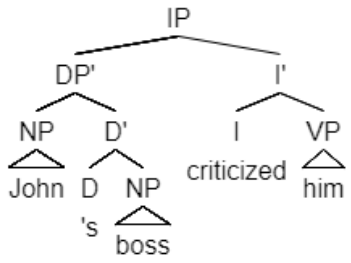

Figure 2: Lack of c-command in binding

On the one hand, pronouns **can** have bound antecedents, but they **do not need** them as in (12) and (13). On the other hand, anaphoras (reflexives and reciprocals) **must** have bound antecedents, and they need to be in the same local domain as in (16).

(16) John$_i$ criticized himself$_i$.

(17) An anaphora must be bound in its local domain (Principle A).

Clearly, pronouns cannot be substituted for anaphoras because they are in complementary distribution.

(18) *John$_i$ said Mary criticized himself$_i$.

Finally, Chomsky postulated the Binding Principles as the following:

(19) Principle A: an anaphora must be bound in its local domain.

Principle B: a pronoun must be free in its local domain.

Principle C: an R-expression must be free.

2.2.2.1 Experimental evidence of the Standard Binding Theory in Brazilian Portuguese

Oliveira et al. (2012) conducted a self-paced reading experiment with the purpose of verifying whether reflexives must be bound locally in Brazilian Portuguese. They manipulated the gender of the local and non-local antecedent candidates and the gender of the reflexives. The authors predicted faster anaphora resolution in cases in which the reflexives would agree in gender with local antecedents, in congruence with the binding theory. The result reported by Oliveira et al. (2012) corroborated their predictions, that is, sentences in which the reflexives agreed in gender with a local antecedent had faster reading times at the reflexives and at the segments following the reflexives. Moreover, it should be mentioned that non-local candidates were not taken into account by reflexive in neither on-line nor off-line measures (comprehension questions). The authors concluded Principle A has proved to be psychologically real in Brazilian Portuguese, since it constrained the reflexive antecedent candidates.

Maia et al. (2012) were interested in comparing pronouns to conceptual pronominal anaphors in Brazilian Portuguese. Conceptual anaphors are those that despite not agreeing morphologically to their antecedents, they are bound to them by a conceptual relation. For example, antecedents like “team” can be retrieved by pronouns like “they”. The authors conducted a self-paced reading experiment aiming at detecting whether Principle B rules both pronoun processing and conceptual

pronominal anaphors. The results corroborated the authors' predictions. Reading times at the pronoun regions were faster when the pronouns were not c-commanded by local antecedents, that is, when they were locally free. In addition, it was not found any difference between pronouns and conceptual anaphors, they both seem to be ruled by Principle B.

Maia et al. (2012) were also concerned about testing Principle C in Brazilian Portuguese. They conducted a self-paced reading experiment with backwards anaphors, in which the pronouns or conceptual anaphors could be bound or not to their antecedents. In case the anaphors were not bound, they would be considered as R-expressions, ruled by Principle C. However, in case the anaphors were bound, Principle B would govern them. The results found by the authors indicate faster reading times at the anaphors regions when antecedent candidates could bind backwards anaphors (Principle B) in comparison to the sentences where the antecedents were free (Principle C). It was not found any difference between pronouns and conceptual anaphors.

Lessa (2014) also found out a preference of Principle B over Principle C in Brazilian Portuguese. Her results showed participants preferred to bind pronouns in backwards anaphors than leave them unbound as R-expressions. Reinhart & Reuland (1993) were also concerned whether pronouns are referring expressions or referential free expressions (R-expressions), as one can see in the next section.

2.2.2 Predicate-Based Binding Theories

Although, Chomsky redefined "The binding theory" by relativizing the size of the locality domain and including the requirement of a subject, this system still has

flaws.

Kuno (1987), from a functional perspective, argued that non-structural factors should be also applied to syntactic analysis. He expanded the vision of the anaphors binding domain including semantic (pragmatic/discourse) factors as, for example, discourse perceptiveness, discourse empathy, definiteness, and the flow of information in the discourse, which includes factors like topic, comment, presupposition, and new and given information.

Kuno (1987) used the term "logophors" to refer to a specific type of reference that takes place when a reflexive should be bound to a local domain that corresponds to a complement clause.

(20) $John_i$ told Mary that [there was a picture of $himself_i$ in the morning paper].

As can be seen in (24), the reflexive "himself" is structurally licensed to be co-indexed with the antecedent "John", in spite of the fact that the antecedent "John" is not in the reflexive local domain. This kind of co-indexation could have violated Principle A; however, "logophors" are often assumed to be exempt from Principle A. This case of "logophors" is also known as picture noun phrases (PNP), which are NPs headed by a representational nouns such as *picture*, *story*, *opinion*, etc. PNPs were the object of study of theoretical linguists such as Pollard and Sag (1992) and Reinhart & Reuland (1993).

Chomsky's Binding Theory was concerned in defining locality domains based on the complementary distribution of anaphors and pronouns. Pollard and Sag (1992, 1994) and Reinhart & Reuland (1993) criticized this approach because anaphors and

pronouns are not in complementary distribution. Instead, they argued that binding is a result of predicate contexts.

Pollard and Sag (1992) and Reinhart & Reuland (1993) focused on detecting either anaphoras and pronouns overlapping or locally unbound uses of reflexives anaphors. The former elaborated a new binding model to be applied to English, while the latter were concerned about Dutch.

Pollard & Sag (1992) developed a model called "Head-Driven Phrase Grammar" (HPSG) with the purpose of abstractly describing and defining the grammar of any language. The HPSG is based on grammatical hierarchies that imply certain constraints to the linguistic objects, which are all described as features structures. One of key concepts of HPSG is "valence", which embodies the notion of argument structure. Thus, words are assumed to select their subjects and complements via "valence features". This applies not only to verbs, but also to all argument-taking words such as nouns and prepositions, for example. The valents of a word have an abstract linear order called obliqueness order, which may differ from the temporal order of their phonetic realizations as in (21).

(21) Argument obliqueness order: subject > object > second object > oblique PP > verb and predicate complements

In HPSG, all NPs and case marked PPs have an index, including phonetically inaudible elements. Each indexed element belongs to one of the three reference types: (a) r-pronouns, which include overt reflexives and the reciprocal "each other"; (b) p-pronouns, which include the non-reflexive ordinary definite pronouns; and (c) non-pronouns, which are all overt indexed NPs, including names

and relative and interrogative pronouns.

From the obliqueness order, the notion of local o-command is defined in (22).

(22) for two indexed valents X and Y of a word, X locally o-command Y just in case it precedes Y in the obliqueness order of that word's valents.

The main difference between Chomsky's Binding Theory and HPSG is the following: in HPSG, not all anaphors in English need to be locally bound as "himself" in (23), but only when they are preceded by other arguments. In other words, local binding is only required when the anaphora is in the same syntactic argument structure as its binder, as in (24).

(23) *Gob said that the newspaper article deeply embarrassed himself.

(24) Gob said that the newspaper article included deeply embarrassing pictures of himself.

Another important difference between Chomsky's Binding Theory and HPSG regards embedded non-finite clauses as in (25). On the one hand, Chomsky's Binding Theory defends that "each other" is bound to "They" as both are in the same local domain; on the other hand, Pollard & Sag (1992) argues that "each other" is critically free once the subject of the embedded clause is the least oblique referent of the embedded verb's sub-categorization frame. They explained that reference to the matrix subject is "accidental" co-indexation.

(25) They would prefer for each other to win.

Pollard and Sag (1992) also defended the idea that discursive factors can also influence binding in "exempt anaphors", name used by the authors to refer to free anaphoras. Exempt anaphors preferentially refer to the "perspective holder" of a discourse as in (26). This way, since the experiencer perspective is more salient, it influences binding.

(26) John was going to get even with Mary. That picture of himself in the paper would really annoy her.

(27)*Mary was taken aback by all the attention John was receiving. That picture of himself in the paper would really annoy her.

Pollard & Sag (1992), reformulated the Principles A and B of the binding theory:

(28) Principle A: (i) Every locally o-commanded r-pronoun is co-indexed with one of its local o-commanders, if one exists; (ii) or interpreted according to certain pragmatic constraints involving logophoricity, contrastiveness, or discourse prominence

(29) Principle B: Every p-pronoun is co-indexed with none of its local o-commanders.

According to Reuland (2006), co-reference is not the only way in which the interpretation of two elements can be related. In (30), "no one" does not refer to an individual; consequently, "he" cannot refer to that individual. Nevertheless, "he" depends on the interpretation of "no one". This within sentence linguistically encoded

dependency is called binding. For Reuland (2006) binding and coreference are distinct processes: binding is an interpretative dependency that is linguistically encoded within sentence grammar, while coreference takes place beyond the sentence scope and does not require syntactic binding.

(30) No one believes he is guilty.

Similarly, Reinhart & Reuland (1993) claimed that coreference is not directly governed by the binding theory, or by any other sentence-level conditions, but falls, together with many problems of anaphora resolution, under discourse theory. Sentence-level coindexation is not relevant for coreference, as it can be obtained only when a pronoun or an anaphora is not coindexed with an antecedent.

Reuland (2006) affirmed that some languages such as Dutch have a richer anaphoric system than English. In Dutch, there are 3 referring devices: *zichzelf*, which resembles the reflexives in English, and it must be locally bound; *hem*, which resembles the pronouns in English, and it must be locally free; and *zich*, which usually cannot be locally bound, but sometimes can, that is, it overlaps between the other two categories.

Reinhart & Reuland (1993) divided the anaphoras into two categories: SELF, which comprises universally local expressions (the English "himself", the Dutch *zichzelf*, the Norwegian *seg selv*, etc.), and SE, which comprises long-distance universally simplex expressions (the Dutch *zich*, the Norwegian *seg*, the Italian *sè*, etc.).

As exemplified in the examples below, in (31) there are two co-indexed arguments, and the SELF anaphora *zichzelf* is one of them. In (32), there is a long

distance dependency; therefore, both SE anaphora *zich* and the pronoun *hem* are possible. In (33), there is an intrinsically (lexically) reflexive predicate, which allows the local SE *zich*.

(31) Jan haat zichzelf / *hem / *zich.

Jan hates himself / *him / *SE.

(32) Jan zag [jou achter zich / hem staan.

Jan saw you stand behind SE / him

(33) Max shaaamt zich.

Max is ashamed.

Reinhart & Reuland (1993) developed the following typology of anaphoric expressions:

	SELF	SE	PRONOUN
Reflexivizing function	+	-	-
R(eferential independence)	-	-	+

Table 1: Anaphoric expression typology by Reinhart & Reuland (1993)

Reinhart & Reuland (1993) argued that much of the complexity of binding systems such as in Dutch results from the interaction between binding and the properties of predicates. Thus, they postulated the "Reflexivity Theory" as the following:

(34) Reflexivity

a) Reflexive predicate: a predicate P is reflexive iff two of its arguments are coindexed

b) Reflexive marking: a predicate P is reflexive marked iff either P is inherently reflexive, or one of P's arguments is a SELF-anaphor

(35) Binding Theory

a) A reflexive-marked predicate is reflexive

b) A reflexive predicate is reflexive marked

In Reinhart & Reuland (1993), Condition A is related to SELF-anaphors and syntactic reflexivity, while Condition B to SE anaphors and semantic reflexivity.

According to Reinhart & Reuland (1993), SELF-anaphors can also be used logophorically in cases in which it does not occupy an argument place of a predicate. Logophoric uses also include the emphatic uses of reflexives as in (36) and in (37).

(36) It angered him that she tried to attract a man like himself.

(37) This letter was addressed only to myself.

The theory developed by Reinhart & Reuland (1993) only encompassed anaphoras, excluding pronouns, because, according to the authors, pronouns are referentially independent since they can be used to select a value directly in the discourse. For Reinhart & Reuland (1993), pronouns should be included in the same category of full NPs, that is, in R-expressions. Therefore, they should be computed by another grammar module distinct from reflexives.

The concept of A-chain (Chomsky, 1986) is closely related to binding. An A-chain is any sequence of coindexation that is headed by an A-position and satisfies the antecedent local domain. This way, each coindexed link, except for the head is c-

commanded by another link, and there is no barrier between any two of the links. From the traditional A-chain definition, Reinhart & Reuland (1993) reformulated the concept of A-chain in their theory.

(38) Condition on A-chains: A maximal A-chain (α_1, ..., α_n) contains exactly one link-α_1-that is both +R and Case-marked.

Based on the condition above, (39) is grammatical because it "She" is a case-marked +R expression, as a result, it can head the A-chain and be coindexed with "herself"; on the contrary, in (40), since "Herself" is neither case-marked nor a +R expression, it cannot head the A-chain, and; consequently, cannot be coindexed with "her". Reinhart & Reuland (1993) emphasized the fact that it seems that any language in the world has nominative case-marking for reflexive anaphors, even highly case inflected languages such as Russian and Icelandic, which corroborates in favor of the idea that SELF anaphors are –R expressions, and cannot head A-chains.

(39) She praised herself.

(40) *Herself praised her.

Bruening (2006) pointed out that "predicate-based theory" such as Pollard & Sag (1992, 1994) and Reinhart and Reuland (1993) has some pros such as: (a) it derives locality conditions on pronouns and anaphors, (b) predicts that binding conditions govern only anaphors and pronouns in argument positions, (c) include exempt anaphors in their account; (d) incorporates a wider range of anaphoric elements than just the two-way split between anaphors and pronouns. However, some

cons are the following: (a) it is not easy to do in traditional grammar; (b) it has troubles to explain reciprocals; (c) it cannot account for anaphors as embedded subjects; (d) it cannot delimit the class of exempt anaphors.

Bruening (2006) highlighted that "Predicate-based Theories" seem to be correct that the defining property of an anaphoras the need for a local antecedent, and that the traditional Condition B that rules pronouns should be put in a different module of grammar.

2.2.3 Summary

In this section Chomsky's Standard Binding Theory (1981, 1986, 1993) was reviewed as well as the Pollard and Sag (1992, 1994) and Reinhart & Reuland (1993) Predicate-based Theories. It seems that none of these three theories is better than the others; they all have flaws.

On the one hand, Chomsky's Standard Binding Theory is concerned in delimiting the local domain of reflexives and pronouns and in which conditions they should bind their antecedents. Chomsky's Standard Binding Theory presupposed anaphoras and pronouns were in complimentary distribution: anaphoras are mandatorily bound to their antecedents in the local domain, while pronouns are optionally bound to antecedents out their local domain. And R-expressions are never bound. Chomsky's Standard Theory is not able to explain the contexts in which anaphors and pronouns overlap, or the contexts in which the conditions seemed to be violated. Moreover, since it is a two-split system (anaphoras and pronouns), it does not contemplate languages such as Dutch, which is a language with 3 types of referring devices.

On the other hand, Pollard & Sag (1992) and Reuland and Reinhart (1993) theories analyzed binding in terms of co-arguments and predicates. Pollard & Sag (1992) and the HPSG theory conditioned binding to the o(blique) command order, eliminating c-command, while Reihart and Reuland (1993) focused on elaborating a theory for reflexive anaphors, banning pronouns from their analysis. According to Reihart and Reuland (1993), pronouns should be considered R-expressions together with full NPs. The Predicate-based Theories of Pollard & Sag (1992) and Reuland and Reinhart (1993) were able to handle the cases of exempt anaphors, and the in the case of Reuland and Reinhart (1993), the three-split system of Dutch.

With respect to reflexives, it seems the three theories reviewed in the section are in agreement with the fact that direct object reflexive anaphors must be bound to the subjects. With respect to pronouns, the case is that this category still vaguely defined in the three theories. Similarly, Chomsky's Standard Theory and Pollard & Sag (1992) stated pronouns must not be bound within their domain and to their complements respectively, while Reinhart & Reuland (1993) preferred to exclude pronouns from binding theory under the reason they are case marked +R expressions, and, consequently, referentially independent.

Finally, it should be highlighted despite its flaws; the Standard Binding Theory is still the most popular binding theory in the literature. A myriad of linguists from all around the world have based their research on it and a lot of psycholinguists have been trying to show its psychological reality in a variety of languages. Since the object of study of this dissertation is Brazilian Portuguese, some studies on binding processing in this language were briefly addressed in this section. They all proved the Standard Binding Theory seemed adequate to explain their results.

2.3 Pronouns and discourse

Reinhart & Reuland (1993) emphasized the importance of distinguishing binding from coreference. While binding is a syntactic dependency that occurs within sentences (intra-sentential), coreference is a discursive dependency that occurs between sentences (inter-sentential). Moreover, these authors argued that pronouns are not anaphors because they are referentially independent; therefore, they should be considered as R-expressions together with full NPs.

Contrary to Reinhart & Reuland (1993), Gordon & Hendrick (1998) proposed a model that approximated intra-sentential and inter-sentential coreference and that differentiated pronouns from full NPs. In addition, Gordon & Hendrick (1998) tried to accommodate interpretative rules from formal semantics and syntactic representations onto discourse representations. They also compromised off-line results from intuitive judgments of grammaticality to on-line measures of language processing.

The Discursive Prominence Theory (Gordon & Hendrick, 1998) describes the representation and processing of reference and coreference in natural languages. This model is based on a tripod:

(i) Form of referring expressions;

The primary function of pronouns is to refer to entities that have been already mentioned in a discourse, while names (or other unreduced referring expressions) introduce new entities into a discursive model. In other words, pronouns are a natural vehicle for coreference, in contrast to names. Repeated names as in (41) are more costly processed than pronouns as in (42), because there will be a moment in the semantic denotation in which “Jane” in the subject position and “Jane” in the object

position would be different referents. In addition, (41) has a longer semantic denotation than (42). This is called the Repeated Names Penalty.

(41) Jane$_i$ thinks she$_i$ is sick.

x
Jane (x)
x thinks x is sick

(42) Jane$_i$ thinks Jane$_i$ is sick.

x y
Jane (x)
Jane (y)
x thinks x is sick

(ii) How syntactic and sequential structure of language is related to discourse models;

The syntactic and sequential structure of a sentence is the major determinant of the prominence of a discourse entity. Syntactically prominent of an NP is related to its height in a syntactic tree and therefore inversely related to its depth of embeddedness. High prominence of a discursive entity affects the interpretation of subsequent referential expressions by facilitating coreference by pronouns and resisting coreference by repeated names (Gordon et al., 1993; Gordon et al., 1999, among others).

(iii) Incremental construction of meaning in discourse.

Each utterance in a discourse adds conditions to the discourse model that serve to further specify the meaning of the model. By the time pronouns are encountered,

there is an immediately search to identity its discursive referent that already exists in the in the discursive model. The suitable antecedent is determined by the "grammatical" features of the pronoun: number, gender, animacy, and reflexivity. If no suitable antecedent is found, an instruction is executed to posit a new discourse referent. An exception to the incremental processing is the backwards anaphora: the pronoun precedes the name with which it corefers. It should be mentioned there is not any repeated name penalty in backwards anaphora with fronted adjuncts. This means that pronouns might not immediately be interpreted in a coreferential fashion when they are contained in fronted adjuncts. This explains why Reinhart & Reuland (1993) considered pronouns as R-expressions.

Moreover, Discursive Prominence Theory also defends pronouns are interpreted based on the semantic plausibility of the event described. For example, in (43), "He" corefers to "Bill", despite "John" being the most prominent discursive candidate, it is not the adequate antecedent; therefore, an additional cost is caused. It should be mentioned that no repeated name penalty was found in this case, which means that elevated reading times did not result from semantic complexity of the passage but rather from difficulty in interpreting the pronoun.

(43) John sent a package to $Bill_i$. He_i received it two days later.

Cases like (43) are called transfer events. Events are characterized by its thematic roles: the verb "sent" assigns the thematic role of source to "John", the thematic role of theme to "package", and the thematic role of "goal" to "Bill". The verb "received" in the second clause also assigns the thematic role of "goal" to "He";

therefore, “He” needs to be reinterpreted as referring to “Bill” instead of the most prominent candidate “John”.

To conclude, the Discursive Prominent Theory (Gordon & Hendrick, 1998) provided integration between generative linguistics, psycholinguistics, computational linguistics, and formal semantics. It is based on the fact that pronouns interpretation relies on discursive context.

Hobbs (1979) highlighted discourse tends to be coherent because it facilitates comprehension and enhances understanding. Consequently, successive sentential units can refer to the same discursive entities, which is known as coreference. This way, coreference is due in part to coherence. Because speakers know discourse is coherent, and they know listeners know this and that they do the best they can to recognize coherence, speakers can leave many entities unmentioned or minimally described in discourse. Hobbs (1979) argued coherence and coreference problems are solved simultaneously by the means of conversational implicatures. According to him, pronoun interpretation utilizes world knowledge during inferential establishment of discourse coherence relations.

Grosz et al. (1995) was also concerned about the relationship between coreference and coherence. According to them, discourse is more than a sequence of utterances; discursive utterances need to be coherent. In order to analyze local coherence of discourse, Grosz et al. (1995) proposed the Centering Theory, which relates focus of attention, choice of referring expression, and perceived coherence of utterances within a discourse segment. According to this theory, certain entities mentioned in an utterance are more central than others, and that this property imposes constraints on a speaker’s use of different type of referring expressions such as full NPs and pronouns. According to the authors, discourse coherence depends on the

compatibility between centering properties of an utterance and choice of referring expressions.

A key concept in Centering Theory is focus, which can be modeled at any point in the discourse by attentional states. In order to avoid confusion with previous uses of the term "focus" in linguistics, the notions of forward-looking (potential discourse focus) and backward looking (current discursive focus) were used by the authors. Grosz et al. (1995) emphasized that one feature that distinguishes Centering Theory from other discursive approaches is the fact that centers of an utterance in general, and more specifically the backward-looking center, are determined by a combination of syntactic, semantic, discourse, and intentional factors. In addition, Grosz et (1995) claimed the ranking of the forward-looking center can be determined by the grammatical role of the entities in discourse: Subject > Object > Other.

The authors investigated how linguistic and attentional states contribute to coherence among utterances within a discourse segment by postulating two rules in the Centering Theory.

(44) Rule 1: If any element of the backward-looking center is realized by a pronoun, then the forward-looking center must be realized by a pronoun also.

Rule 1 reflects one function of the pronominal reference: the use of a pronoun to signal the hearer that the speaker is continuing to talk about the same thing. An example of Rule 1 breaking is in (45).

In (45), the use of the pronoun to refer to Tony in utterance (e) may confuse the hearer. From utterances (a) to (d), Terry has been the center of attention, and the

most likely referent of “he” in (e). It is only when the word “sick” is heard, that it is clear that the pronoun “he” in (e) refers to Tony and not Terry.

(45) a. Terry really goofs sometimes.

b. Yesterday was a beautiful day and he was excited about trying out his new sailboat.

c. He wanted Tony to join him on a sailing expedition.

d. He called him at 6 am.

e) He was sick and furious at being woken up so early.

(46) Rule 2: Sequences of continuation are preferred over sequences of retaining; and sequences of retaining are to be preferred over sequences of shifting.

Rule 2 means that frequent shifting leads to a lack of coherence, and that the basis for a local coherence is the continuation of the center and the use of retentions when possible to provide smooth transitions a new center. An example of the use of different types of transitions is in (47).

(47) a. John has been having a lot of trouble arranging his vacation.

b. He cannot find anyone to take over his responsibilities. (he=John)

c. He called up Mike yesterday to work out a plan. (he=John, CONTINUE)

d. Mike has annoyed him a lot recently. (he=John, RETAIN)

e. He called John at 5 am on Friday last week. (he=Mike, SHIFT)

Grosz et al. (1995) also addressed definite noun phrase interpretation. The authors highlighted no matter rich a model of context is, it will not be possible to fully constrain the interpretation of an utterance when it occurs. Thus, the authors are in favor or a partial interpretation computation that refines that interpretation on the basis of subsequent utterances. An example is that there are several interpretations of the NP "the Vice-president of the United States". In (48), the interpretation in (a) provides the basis for the interpretation of "he" in (b); however, this is not the case of "he" in (c).

(48) a. The Vice-President of the United States is also President of the Senate.

b. Right now, he is the President's key person in negotiations with Congress.

c. As ambassor to China, he handled many tricky negotiations, so he does well this job.

To conclude, Centering Theory contributed to understanding the discursive status of pronouns, and how the choice of referring expressions can model attentional state in discourse and maintain local coherence.

The following subsection will review some experimental research on the role of discourse in pronoun processing in Brazilian Portuguese.

2.3.1 Experimental evidence of discursive factors in pronouns processing in Brazilian Portuguese

In this subsection, some experimental works that examined the role of discursive factors in pronouns processing in Brazilian Portuguese will be briefly reviewed.

These studies focused on investigating grammatical role parallelism, repeated names penalty, and the psychological reality of different coreference forms such as full and null pronouns.

According to the Centering Theory (Grosz et al., 1995), parallelism would only influence subject referents, that is, pronouns in either subject or object positions would tend to refer to referents in the subject position. However, parallelism can be extended to object positions as well. In psycholinguistic literature, parallelism between pronouns and their antecedents can facilitate antecedent selection. For example, a pronoun that occupies a subject position tend to refer to an antecedent that also occupies the subject position, while a pronoun that occupies an object position tend to refer to an antecedent that also occupies the object position.

Corrêa (1998) conducted an off-line questionnaire with the purpose of checking whether parallelism of grammatical roles would influence full and null subject pronouns in Brazilian Portuguese. The author manipulated the type of clause, which could be independent, coordinate, or subordinate. The results suggested evidence in favor of the influence of parallelism in pronoun resolution, that is, there is a preference for full and null subject pronouns to retrieve antecedents in the subject position. Moreover, null pronouns were preferred over full pronouns especially in coordinate and subordinate clauses, which contain a certain level of syntactic dependency.

Leitão (2010) conducted a self-paced reading experiment in order to examine the relationship between animacy and parallelism in coordinate clauses with object pronouns in Brazilian Portuguese. He predicted that inanimate antecedents would be more costly retrieved than animate antecedents, respecting the object parallelism. The

results corroborated this prediction, object pronouns were more easily processed when their referents were animate and occupied the object position in the sentences.

In contrast, Hora (2014) did not find out any effects of grammatical role parallelism for full pronouns in her self-paced reading experiments, that is, full subject pronouns could equally refer to antecedents in either the subject or in the object position. However, she found out null subject pronouns prefer to refer to antecedents in the subject position. Finally, her results indicated coreference is facilitated with full pronouns rather than null pronouns. Similar results were found by Machado (2016).

Leitão (2005) was the first to study the repeated names penalty in Brazilian Portuguese. He conducted a self-paced reading experiment comparing full and null object pronouns in coordinate clauses. The results indicated coreference was more difficult to be processed with repeated names than with pronouns, that is, pronouns rather than repeated names are a more natural way of expressing coreference. Furthermore, Leitão (2005) found out that object pronouns tend to refer to antecedents in the object position, which is evidence in favor of the parallelism.

Leitão et al. (2012) were interested in investigating the repeated names penalty in Brazilian Portuguese with subject and object pronouns. They conducted an eye-tracking experiment with the purpose of confirming Queiroz & Leitão (2008) results in a self-paced reading experiment. The results found by Leitão et al. (2012) corroborated Queiroz & Leitão (2008). In other words, pronouns are more easily processed than repeated names in either subject or object positions.

The conclusion is grammatical role parallelism in pronoun resolution is still questionable. Some studies showed results in favor of parallelism, while others did not. In the section “Interpreting the pronouns”, Kehler (2007) will propose parallelism

per se does not exist, but it is an epiphenomenon of coherence. With respect to repeated names penalty, it was found evidences that corroborate the fact that pronouns rather than full NPs are a more natural way of expressing coreference.

2.3.2 Summary

This section focused on presenting the three major works that addressed the relationship between discourse and pronouns resolution: Hobbs (1979), Grosz et al. (1995), and Gordon & Hendrick (1998).

Hobbs (1979) was the first to relate pronouns and discourse. According to him, pronoun resolution is influenced by coherence factors based on world knowledge inferences. Grosz et al. (1995) developed the Centrality Theory, which postulates pronoun resolution is affected by the topichood conveyed in discourse. Gordon & Hendrick (1998) elaborated the Discursive Prominence Theory, which proposed pronouns rather than repeated full NPs are a more natural way of expressing coreference. They also claimed that pronouns tend to refer to the more prominent referent in discourse, which in the majority of the cases is the subject. Gordon & Hendrick (1998) argued discursive information is incrementally constructed in the mental model; this way, factors such as plausibility need to be taken into account.

Experimental data in Brazilian Portuguese concerning the role of discourse in pronominal resolution still needs more investigation. Questions such as grammatical role parallelism still needs more attention since the results found by psycholinguists seem contradicting. Nevertheless, it seems the Discursive Prominence Theory can explain the fact that null pronouns tend to refer to antecedents in the subject position (Hora, 2014; Machado, 2016). This can be explained by the fact that since null

pronouns lack phonological realization they are less salient in memory; consequently, they need to refer a highly salient entity in discourse, which is the subject. With respect to the repeated name penalty, it seems pronouns are preferred over repeated names in coreference processing.

2.4 Discourse and gender in pronouns processing

In this section, some experimental studies interested in examining the role of discourse and gender cues in pronouns processing will be reviewed.

Arnold et al. (2000) investigated how gender information and accessibility influence the initial processes of pronoun interpretation. The authors explained that it appears pronoun comprehension is influenced by the accessibility of potential referents. For example, in (49), Clinton is more accessible than Gore because Clinton is coming first in the sentence. Thus, comprehenders might have difficulty assigning the pronoun "he" to the correct referent, Gore. However, in (50), it should be easier to understand the pronoun "she", because it has only one referent that matches the gender of the pronoun.

(49) Clinton confessed to Gore when he asked for the truth.

(50) Clinton confessed to Albright when she asked for the truth.

Arnold et al. (2000) monitored the participants' eye movements while they viewed a picture and listened to a text describing the picture. Each text had four clauses, broken into two sentences. In Experiment 1, the first clause mentioned the two characters; the second mentioned some other object in the picture. The third clause began with a pronoun referring to one character or the other, and the final

clause provided concluding information without mentioning either character individually.

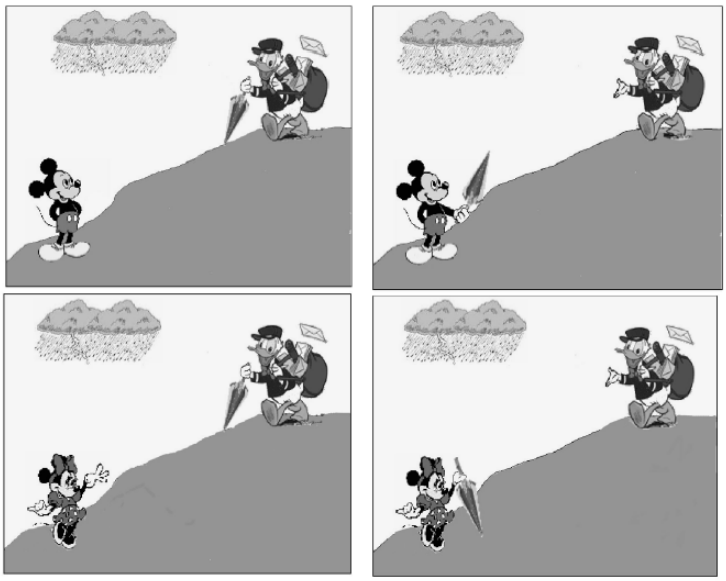

Figure 3: Sample of the materials used by Arnold et al. (2000)

(51) Donald is bringing some mail to {Mickey/Minnie} while some violent storm is beginning. He's/ She's carrying an umbrella, and it looks like they're both going to need it.

The results found by Arnold et al. (2000) showed listeners are able to rapidly converge on the intended referent based on-order-of-mention and gender cues. Only when both of these cues were unhelpful (same gender/ 2nd-mention condition) listeners showed difficulty identifying the pronoun referent.

In Experiment 2, the text was changed to increase accessibility of the 1st-mentioned character and making the 2nd mentioned character inaccessible by adding a pronominal reference to the first character.

(52) Donald is bringing some mail to {Mickey/Minnie}. He's sauntering down the hill, while some violent storm is beginning. He's/She's carrying an umbrella, and it looks like they're both going to need it.

The results of Experiment 2 replicated Experiment 1. Early target identification occurred when either gender or accessibility information was present. Since pronominal reference to the second character in the same gender condition was highly infelicitous, subjects were often garden-pathed in the same-gender/2nd mention condition.

Arnold et al. (2000) concluded gender and accessibility affect the initial processes of pronoun resolution in both experiments. When either factor signaled the correct pronoun referent, participants' eye movements indicated they began considering the target referent more than the competitor referent soon after the offset of the pronoun. When neither gender nor accessibility was sufficient to resolve the pronoun, participants did not rapidly converge on an interpretation of the pronoun. In addition, it seems that gender and accessibility information influence referent consideration at the same moment, therefore, this study support a dynamic model of language processing, where multiple sources of information are used probabilistically to guide referential processing.

Foraker & McElree (2007) argued pronoun resolution requires access to previously processed representations. The antecedent representation that a pronoun refers back to has to be accessed in the comprehender's discourse model, and then aligned and integrated with the pronoun so that a coherent representation of the text or conversation can be formed. In (53), a representation of the antecedent "toothbrush"

must be made available in the comprehender's mind and integrated with the pronoun "it".

(53) Where's my toothbrush? Have you seen it?

Foraker & McElree (2007) claimed pronouns are used to refer to psychologically more prominent entities, while forms with more semantic information are preferred for less prominent ones. Pronouns such as "it" tend to refer back to concepts that are highly predictable or salient in discourse situation, such as "toothbrush". For less prominent concepts such as in (54), it is necessary to use more lexically specific forms than a pronoun.

(54) Where's my black toiletries kit with the toothbrush and toothpaste in it? Have you seen my toothbrush?

Foraker & McElree (2007) showed that studies of memory retrieval indicate that information that is actively maintained in focal attention can be faster accessed than information stored in a more passive memory state. One of the ways to increase prominence is through focus. Clefting can preferentially direct a reader's or listener's attention to the clefted entity. This way, pronominal coreference with a clefted antecedent appears to be easier than coreference with a nonclefted antecedent. Taking this into account, Foraker & McElree (2007) conducted two speed-accuracy tradeoff (SAT) experiments in which participants were presented to the fragmented stimuli and were asked to press a button indicating acceptable or unacceptable judgments.

The authors also conducted two eye-tracking experiments with the purpose of capturing early processing effects.

The results found by Foraker & McElree (2007) indicated that syntactic clefting did not affect the speed of accessing an antecedent representation, which contradicts claims that discourse-focused information is actively maintained in focal attention. Rather, clefting simply increased the likehood that an antecedent representation was successfully retrieved, which suggests that clefting only increases the strength of a representation in memory. Similarly, eye-tracking measures showed that clefting did not affect early bonding of the pronoun and the antecedent, but did ease later integration. The authors explained clefting made an antecedent representation more available in memory. Clefting does not seem to alter the retrieval cues provided by the pronoun, but it seems to affect how the antecedent was first encoded in memory. Furthermore, conditions with a gendered "he" or "she" pronouns were faster and more likely to be successfully resolved than those of an "it" pronoun. Gendered pronouns facilitated resolution because they are less ambiguous and did not lead to much competition between antecedents or required reanalysis as often as "it"-pronouns did. This way, both clefting and pronouns increased the availability, or distinctiveness, of an antecedent representation in working memory.

Several studies on pronoun resolution and on corpus analysis show pronouns tend to refer to the higher accessible entity in the discourse, that is, the antecedent generally occupies the focus of attention in natural discourse. This way, according to Greene et al. (1992), the most relevant question about pronoun processing is to find out how discursive processing makes the referent highly accessible, rather than to find out how pronouns find their referents. Greene et al. (1992) defended discursive top-down processes are more important than bottom-up information such as gender. They

stated gender is not automatically used to select a referent from equally accessible referents.

Rigalleau et al. (2004) was concerned in investigating the automatic processing of gender information in pronouns in French. They conducted clause-by-clause self-paced reading in Experiments 1 and 2, and probe recognition tasks in Experiments 3 and 4. The participants were asked to read sentences where a pronoun appeared in a subordinate causal clause. They varied the implicit causality of the verb in the subordinate clause, the congruity of the subordinate clause relative to the bias of the verb, and the relevance of the pronominal gender cue.

Rigalleau et al. (2004) found out the reading time of the pronominal clause was shorter when pronominal gender was sufficient to resolve the pronoun than when pronominal gender was in agreement with two nouns mentioned in the main clause. They authors explained that as the participants knew they would be questioned about the pronoun, it would be necessary to reduce the level of activation of the nonreferent. Thus they strategically used gender information to concentrate the integration process on just one entity. However, when the participants were not asked about the pronouns, the relevant gender cue did not significantly accelerate the reading time of the subordinate clause. Nevertheless, participants were still sensitive to a gender disagreement between the pronoun and the two potential referents mentioned in the previous clause. Therefore, even in a condition where they were not supposed to strategic use the gender cue to select the referent, the participants remained sensitive to pronominal gender. This means that pronominal gender processing seems automatic.

Rigalleau et al. (2004) was also interested in examining the role of accessibility in pronominal resolution. They used long texts where one of the potential

referents was mentioned 35 words before the pronoun, and the other potential referent was mentioned 6 words before the pronoun. This way, they expected to place the closer referent in the focus of attention by the time the pronoun was encountered. The authors reported that when there was no strategic reading, probe recognition times of potential referents did not seem to be influenced by pronominal processing; however, they detected longer reading times when pronouns disagreed with the most accessible antecedent. This suggests that participants were sensitive to gender disagreement.

Rigalleau et al. (2004) explained that in Italian, which is a pro-drop language, comprehension system is prepared to deal with sentences in which no processing of a pronoun is required when a single entity is in focus. This justifies the results found by Greene et al. (1992). However, in non-pro-drop-languages like English or French, a pronoun must be produced to refer to the most highly accessible referent. Rigalleau et al. (2004) called this "gender coindexation process". They also claimed in languages where gender can be a purely grammatical features, that is, in languages with grammatical gender such as French, gender coindexation process involves purely morphosyntactic features, without using conceptual gender based on the sex of the referents.

Rigalleau et al. (2004) concluded that automatic gender is consistent with the hypothesis that the main function of a pronoun is not to create accessibility but to confirm accessibility.

2.4.1 Summary

Arnold et al. (2000), Foraker & McElree (2007), and Rigalleau et al. (2004) were concerned about the influence of discourse and gender cues in pronoun resolution

processing. Arnold et al. (2000) defended a multi-source probabilistic mechanism for pronouns resolution in which both syntactic cues such as gender and discursive cues such as accessibility (subject topichood) can be taken into account simultaneously.

Foraker & McElree (2007) showed that keeping a referent in memory focus only increases the chances of this antecedent being retrieved (accessibility), but it does not mean it will be retrieved faster (availability). This way, a highly prominent referent does not have any especial cue encoded, but it is the way its cues are encoded that is different from other referents. These authors also found out gendered pronouns like "he" and "she" can be more easily processed than ungendered pronouns like "it". Gendered pronouns are less ambiguous, which reduces the competition among antecedents.

Finally, Rigalleau et al. (2004) was interested in the relationship between accessibility (referent recency) and gender agreement. Their results indicated gender agreement was only used to check for accessibility, and not to create accessibility. Gender agreement was only helpful in processing in a strategic reading, that is, they readers knew they would be asked about the pronouns. Nevertheless, sentences in which the pronouns did not have any agreeing referents were more difficult to process, which means that readers were sensitive to gender agreement. They also claimed that gender agreement in languages like French is not semantically based, but is a purely grammatical relation.

2.5 Pronoun Interpretation

Kehler (2007) claimed the majority of studies in the literature have assumed a particular type of process underlies pronoun interpretation, which he calls SMASH paradigm: Search, Match, and Select using Heuristics.

(55) a. Search: collect possible referents within some suitable contextual window

b. Match: filter those referents that fail "hard" morphosyntactic constraints such as number, gender, and person agreement, and intra-sentential syntactic binding constraints.

c. Select using Heuristics: select a referent from those that remain by applying a set of heuristically based "soft" preferences such as grammatical role ranking parallelism, among others.

Kehler (2007) explains SMASH procedure differs across different theories and algorithms. However, he argued the SMASH way of explaining pronoun interpretation should be abandoned. Kehler (2007) started showing 3 types of examples in which the Search and Match steps of SMASH paradigm fail to explain coreference.

Firstly, Kehler (2007) pointed out not all entities are salient enough to pronominalize. For example, in (56), "Doberman pinscher" does not license a subsequent pronominal reference despite the fact it is the most recently mentioned entity, occurs only one sentence back, and it is the only entity that satisfies the number restriction on the pronoun.

(56) Two Sears employees delivered some new appliances to my neighbors with the Doberman pinscher.

a. # It's the same dog that bit Susan last summer.

b. That's the same dog that bit Susan last summer.

On the hand, in (56), "He" can felicitously corefers to a referent in the phrase "with the wild child", which occupies the same position as the "Doberman pinscher". The speaker's purpose behind the utterance of "with the Doberman pinscher" is to restrict the reference of the NP it modifies to a unique set of neighbors. Once these neighbors are identified, it appears to have no further contribution to the overall proposition. While "with the wild child" is a conversationally relevant description. That is, this choice may have been intended to create an expectation that the ensuing discourse will address the inadvisability of a couple with a wild child buying expensive china, which gives the wild child a role in the discourse that goes beyond restricting the reference to a unique set of neighbors. This suggests that a discourse process as high-level as reasoning about the intentions that underlie a speaker's choice of linguistic expression is necessarily intertwined with the seemingly lower level process of pronoun interpretation.

(57) Lenox delivered new expensive china to my neighbors with the wild child. He'll have it all broken within a week.

Secondly, Kehler (2007) claimed there is an apparent interference between "soft" preferences and "hard" constraints. Sometimes even referents in the position of subject cannot be felicitously pronominalized in the subsequent utterance. Informants

failed to corefer the pronoun “her” to “Margaret Thatcher” in (58). They reported a felling that the pronoun should refer to “Reagan”, as if the speaker was confused about his gender.

(58) ?? Margaret Thatcher admires Ronald Reagan, and George W. Bush absolutely worships her.

In (58), not only “Margaret Thatcher” not only is in a prominent and salient position as the subject of the sentence, but it is also a plausible referent and it is the only referent that meets the gender restrictions of the pronoun. That is going on in this example is that “soft” preferences for grammatical parallelism trumps “hard” constraints like gender agreement. This means that Match step cannot come strictly before Select step; they would need to be integrated somehow.

Finally, Kehler (2007) discussed conjoined referents. In (59), the pronominal coreference is infelicitous because “Bush” and “Blair” are evoked from equivalent grammatical positions, and presumably have the same degree of salience in discourse. Thus they are indistinguishable as possible referents of the pronoun. On the one hand, the criteria “first mentioning” favors “Bush”; on the other hand, the criteria “recency” favors “Blair”.

(59) ?? Bush and Blair gave a press conference, and a reporter asked him a question.

The change of gender of one of the conjuncts does not improve the felicity of the pronominal reference as in (60).

(60) ?? Rice and Blair gave a press conference and a reporter asked him a rude question.

Kehler (2007) explained there is more to determining the referent of a pronoun than degree of salience, that is, that there is a TOPICHOOD requirement at play. In both (59) and (60), the entities introduced in the conjoined subject NP presumably must either serve as topic together or not at all – it cannot be that one entity is the topic and the other is not.

The second part of Kehler (2007) is addressed to the third step of SMASH. Many approaches to pronoun interpretation encode a preference for a pronoun to refer to the subject of the previous sentence because (a) the subject position accords greater degree of SALIENCE to its occupant than do other grammatical roles; and (b) the subject position is the canonical place from which to mention a discourse TOPIC, which generally is considered to be continued. However, there are a variety of situations in which the subject assignment preference is neutralized. For example, in constructions of transfer of possession as in (61), change of state as in (62), or in causal relations between the clauses as in (63).

(61) John passed the comic book to Bill. He began reading.

(62) Ken admired Geoff. He knows a lot about cars.

(63) Colin pushed Don. He tumbled to the ground.

In the examples above, the referred referent occupies the object position, even though the subject position entity is morphosyntactically compatible with the

pronoun. In these cases, it would appear that semantic and world knowledge considerations are the determining factors.

Many authors have also argued for a preference in pronoun interpretation based on grammatical role parallelism, that is, pronouns will be preferentially associated with an antecedent in a parallel grammatical role.

(64) Margaret Thatcher admires Hillary Clinton, and George W. Bush absolutely worships her.

In (64), "her" is in the object position and, according to parallelism; it corefers to "Hillary Clinton", which is also in the object position. However, world knowledge strongly posits "Thatcher" as the most adequate referent. Parallelism is also put into question by native speakers is several situations reported from Kameyama (1996), as in (65) and in (66).

(65) John kicked Bill. Mary told him to go home. (he = John)

(66) Bill was kicked by John. Mary told him to go home. (he = Bill)

This way, both subject assignment and grammatical role parallelism preferences can be counterexemplified. Kehler (2007) stated grammatical role parallelism should be abandoned, since it is an epiphenomenon of coherence.

Kehler (2007) argued the SMASH approach besides presenting several deficiencies, it imposes a large amount of processing that has to occur by the time the pronoun is encountered. He reminded pronouns are supposed to facilitate discourse communication, and not hinder it. Pronouns need to be easier processed than repeated

names, for example. This way, if pronoun interpretation were really hard as the SMASH approach suggests, why would a speaker choose to use it?

Kehler (2007) criticized Hobbs (1979)'s approach in reducing coreference as a form of coherence, since it fails to explain the importance of grammatical form and information structure in pronoun interpretation as in (65) and (66). Kehler (2007) also criticized the Centering Theory (Grosz, 1995) since it does not address the semantic relations and world knowledge inferences necessary in coreference. Thus, Kehler (2007) concluded an adequate coreference analysis must capture the interaction between coherence establishment and information structural constraints on pronouns and their referents.

Kehler and Rhode (2013) proposed reconciliation between these two approaches. Based on the results of a series of psycholinguistics experiments analyzed through a Bayesian probabilistic model, the interpretation bias between referents and pronouns might be determined by:

(67) a. Expectations that comprehenders have in coherence relations, which influences top-down expectations about referent next mention (regardless of the referring expression used);
b. Centering-style constraints on pronoun production, which provide bottom-up evidence about the topichood status of the referents that are specific to the speaker's decision to use a pronoun.

Kehler & Rhode (2013) argued that their model integrates top-down expectations and bottom-up linguistic evidence. This way, by Kehler & Rhode (2013)

compromised both Hobbs (1979)`s coherent bias and the Centering Theory (Grosz et al., 1995)'s topichood bias.

2.5.1 Summary

Kehler (2007) discarded the idea that pronoun resolution is based on the SMASH approach (Search, Match, and Select). This process seems to complex and time consuming for pronouns, which are supposed to facilitate communication. Kehler (2007) showed examples in which the three steps of the SMASH approach seem to be simultaneously interconnected rather than respecting a linear order.

Kehler & Rhode (2013) proposed that pronoun resolution might be based on a probabilistic approach in which world knowledge inferences and questions related to discursive topichood are calculated with the purpose of retrieving the most adequate referent in context. In this view, the key to understand pronoun resolution is the integration between top-down and bottom-up information.

2.6 Memory retrieval

2.6.1 Memory models

Atkinson & Shiffrin developed the most influential computational models of memory in 1968. In this model, memory would be consisted of 3 stores: a sensory register (lasts less than a second), a short-term memory (lasts until 18 seconds) and a long-term memory (unlimited). According to the authors, information linearly passes from one store to the other store, like a computer. The environment would provide the input

to sensory memory, which transfers information to short-term memory. This transference can be impacted by attention problems. Once the information is in short-term memory, it can be either recalled or transferred to long-term memory. In this case, transference relies on rehearsing or repetition, and if either of these does not occur, the information is forgotten or lost due to decay.

Inspired by William James, who affirmed short-term memory (primary memory) is qualitatively different from long-term memory (secondary memory), Baddeley & Hitch (1974) attempted to elaborate a more accurate model than Atkinson & Shiffrin in 1968. In their model, short-term memory, unlike long-term memory, would have specific buffers such as the verbal buffer (where phonological loop in charge of verbal rehearsal) and the visuo-spatial buffer. Besides these two buffers, the authors defended the existence of a central executive buffer, which organizes the interplay between short-term memory buffers and long-term memory. Later, Tulving divided long-term memory into declarative and procedural buffers. The former is related to facts, and it is subdivided into semantic memory (meanings) and episodic memory (events), while the latter is related to knowledge and skills. There have been a myriad of empirical evidences in favor of the dual-store memory model. Studies on brain-injured patients showed dissociation between short-term memory and long-term memory, that is, patients with a deficit in short-term memory and preserved long-term memory and vice versa (Shallice & Warrington 1970, Vallar & Papagno 2002). Furthermore, several psycholinguistic studies corroborated in favor of a dual-store memory model, as for example Sachs (1967).

According to Sachs (1967), when language is comprehended, it seems the meaning of what is heard or read is remembered to some extent, but unless special attention is given to the style or other characteristics of the words, the exact wording

is forgotten. She was interested in how sentences are stored in memory. She hypothesized sentences are encoded only with respect to their meaning. The participants of her study heard discursive passages, and after each passage, they heard a sentence which was either identical to a sentence to a sentence that had occurred in the passage, or was changed in some slightly way. Then the participants were asked to respond "identical" or "changed" rating their confidence on a scale and classifying their change as "meaning" or "form". Her findings indicate recognition for the form of a sentence declines much more rapidly than recognition memory for the meaning. If the test sentence followed the original immediately, recognition was high, but after 80 syllables of discourse interpolated, recognition was close to chance. Once form was stored for a shorter interval than meaning, it seems these two types of information have distinct types of storage: short-term memory and long-term memory respectively.

Alternatively, unitary-store memory models defend there is not a qualitative difference between short-term memory and long-term memory, as, for example, they seem to be affected similarly by medium temporal lobe injuries. Neuropsychological studies showed the medium temporal lobe is involved in both long-term memory (declarative or episodic memory) and short-term memory tasks (novel relations or binding) [Gabrieli et al., 1997; Squire, 1992; Ranganath & Blumenfeld, 2005; Nichols at al, 2006]. Moreover, patients with perissylvian cortical lesions, who were believed to have deficits in only short-term memory, actually show deficits in other areas beyond short-term memory, as general phonological processing. This way, it seems short-term memory and long-term memory are not architecturally separable.

Defenders of the unitary-store model argue short-term memories consist of temporary activations of long-term memory (Anderson, 1983; Cowan, 2001;

McElree, 2001; among others). The idea is that there is only one memory store (long-term memory), which is bi-partite in passive and active memory. The active memory is what is within the focus of attention, and since the focus of attention is very limited, the exchange of information between passive and active memory needs to be fast. And it seems that working memory is the one responsible for this interchange of information between active and passive memories.

Regardless of model approaches, the discussion over the limit of short-term memory (for dual-store models) or focus of attention (for unitary-store models) is crucial. Baddeley et al. (1974) suggested verbal memory is limited to 2 seconds, while Cowan (2001), urged the focus of attention would be of approximately 4 items. However, McElree (1996, 2001) indicated the focus of attention consists of a single item.

According to unitary-store models, memory retrievals are the result of the strength of activation, which varies based on recency and frequency of occurrence. McElree (1996) was especially concerned about the effects of recency in memory retrieval. He reported the results of a speed-accuracy tradeoff on word lists, indicating retrieval accuracy was faster and higher for the most recent item, while all the other items were similarly retrieved independently of their position in the sequence. Thus, it seems that the focus of attention is not related to capacity, but to representation.

2.6.2 Serial or parallel memory activation

McElree (2000) affirmed language comprehension requires constructing a hierarchical representation from a liner sequence of symbols. This task entails resolving dependencies between constituents that can be separated by an indefinite

amount of material. To resolve long-distance dependencies, on-line processes must have access to previously processed constituents. Working memory is the "work space" where products of prior analyses are maintained and eventually accessed and modified.

McElree (2000) urged basic binding operations in comprehension are mediated by memory representations that are content addressable, which means that syntactic and semantic constraints provide direct access to relevant representations without the need of search through irrelevant information. The content-addressable system contrasts with traditional views that defends working memory requires a serial search. McElree et al. (2003) stated direct access and search and mechanisms could be empirically contrasted by examining the effect that interpolated material has on the speed of retrieval. On the one hand, in a serial search mechanism, retrieval speed would slow as potentially interfering material is added to the memory system. On the other hand, additional material would not affect retrieval speed in a memory system with direct access mechanism. However, it is a fact that interpolated material may decrease the quality of a memory representation, making the representation less likely to be recovered in any particular context. Nevertheless, direct access mechanism allows representations of different quality or strengths to be retrieved at equal time.

McElree (2000) and Foraker & McElree (2007) were interested in checking whether the availability and the accessibility of a representation would be affected by the amount of interpolated material. What they meant by availability is the probability of maintaining a representation in memory, while accessibility is the time it takes to retrieve the representation. Both works conducted speed-accuracy tradeoff (SAT) tasks: McElree (2000) tested filler-gap dependencies with long linear distance between targets and gaps, while Foraker & McElree (2007) tested the same type of

dependency but with long hierarchical distance; moreover, they tested nonadjacent subject-verb dependency, and two filler-gap dependency.

Despite the variety of the types of dependencies tested, the results found by McElree (2000) and Foraker & McElree (2007) indicated the availability of the representations decreased as additional material was processed; however, its accessibility was not affected by the amount of interpolated material. The authors concluded that these results corroborate in favor of the idea that sentence comprehension is mediated by a content-addressable memory system.

2.6.3 Content-addressable memory model

Lewis et al. (2006) observed understanding spoken and written language in real time requires the rapid integration of prior linguistic material with present material. Thus, language processing must maintain, at least momentarily, some memory of linguistic material. The authors questioned the processes in which working memory retrieves previous interpreted information and the constraints that may exist on those processes.

Based on a cognitive architecture computational model called Adaptive Character of Thought-Rational (ACT-R) by Anderson et al. (2004) and Anderson & Lebiere (1998), Lewis & Vasishth (2005) and Lewis et al. (2006) proposed a new model capable of explaining the content-addressable memory mechanism. According to this model, prior information that was previously interpreted is retrieved by a parallel search based on a set of grammatical cues generated by a target. This set of retrieval cues consist of several types, including structural, morphological, semantic, and discursive (among others).

According to Lewis & Vasishth (2005) and Lewis et al. (2006), the parallel search in memory can be affected by similarity-based interference and decay factors. Similarity-based interference occurs when the overlap between the items in memory and the retrieval cues increase, reducing the strength of association between the cue and the target item, as a great number of items will be associated with the cue. Consequently, memory failure rates increase, and distractors, that is, candidates that partially-match the cues can sometimes be retrieved. However, decay occurs when reactivation of certain items are eliminated or more difficult to happen. Decay can be a result of, for example, increased intervening material between targets and items. In other words, distant items may decay over time.

The content addressable memory model can also be used to explain how pronouns retrieve their antecedents in memory.

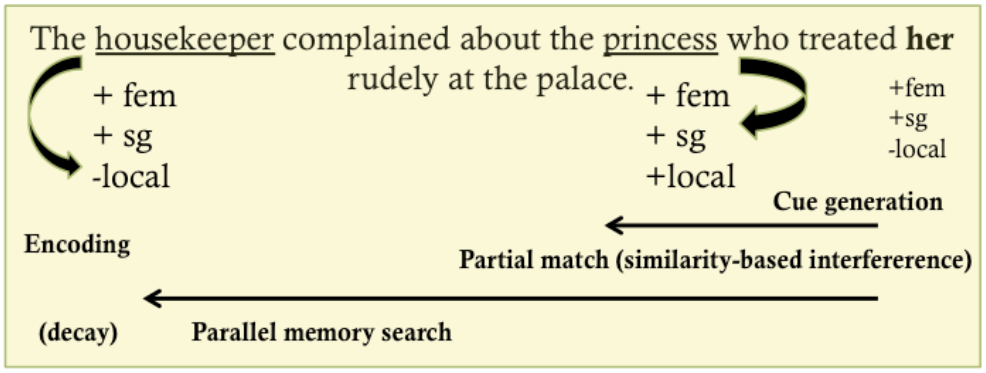

Figure 4: How antecedent retrieval works in content addressable memory. Based on Lewis et al. (2006)

The example illustrated in Figure 4 shows during the encoding phase, all information is interpreted and stored in memory. By the time the pronoun is encountered, a group of grammatical cues is generated is order to retrieve the

antecedent. In the example portrayed in Figure 1, the antecedent must not be local[2], and it must be feminine and singular. After that, there is a parallel search in memory and two candidates that are similar to the cues generated by the target are found: "housekeeper" and "princess". The former candidate is a perfect match; however, although the latter candidate is only a partial match as it is local, it can interfere with memory retrieval, the so-called similarity-based interference effect. Candidates like "princess" are called distractors according to the content addressable model. In addition, in this example, "housekeeper" can also decay over time, as it was stored in memory before "princess", which, in this case, is more recent. Thus, according to this model, distractors such as "princess" can sometimes be erroneously retrieved as antecedents as a result of a failure caused by both similarity-based interference effects and decay factors. In other words, although "princess" is in Principle A domain, it can be considered a potential pronominal antecedent due to the fact it matches the morphological cues generated by the pronoun "her".

It should be mentioned the content addressable memory model does not intend to replace The Binding Theory, but it aims to explain how memory uses content cues to retrieve previous encoded information. This model can be applied to other syntactic dependencies such as subject-verb agreement, for example.

2.6.4 Extended content-addressable memory model

[2] It is important to mention that the status of the [-local] feature can be questioned, as it seems awkward that languages would have this feature specified for each and every item. However, it is assumed that it is actually a relational feature that is only specified in binding dependencies.

Engelmann et al. (2015) realized the literature shows a great variability of results on content-addressable memory, and some of them are not even predicted by the model. Thus, they proposed an extension to the classic content addressable memory model in order to better explain already published results. They reviewed 69 experiments on reflexive-antecedent and subject-verb dependencies and presented the results of a computational model that showed prominence correction and associate cues could successfully predict previous results in the literature.

Engelmann et al. (2015) stated similarity-based interference can cause elevated reading times, which is called "inhibitory interference". Based on Lewis & Vasishth (2005), they explained this inhibitory effect is motivated by a competition between the target and the distractor. Since the amount of activation associated with a retrieval cue is shared between all matching items, the presence of competitors in memory will reduce each item's activation. Since retrieval speed is a function of an item's activation, reduced activation due to cue-matching distractor results in longer retrieval latency as compared to a condition without a cue-matching distractor.

Moreover, Engelmann et al. (2015) pointed out that, according to Lewis & Vasishth (2005), the similarity-based interference increases the probability of erroneously retrieving the partially matching distractor. These occasional misretrievals are predicted to cause incorrectly formed dependencies, affecting comprehension in the respective trials. In special occasions, misretrievals of the distractor can lead to an observed speed-up in mean reading times. This is called "intrusion" and refers to cases where a distractor causes an ungrammatical sentence to be perceived as grammatical, for example.

However, Engelmann et al. (2015) claimed when target and distractor do not overlap in the manipulated feature in the distractor-match condition, no similarity-

based interference is predicted. Nevertheless, because both target and distractor partially match the retrieval cues, the probability of erroneously retrieving the distractor is predicted to increase. This causes shorter retrieval latencies in the distractor-match conditions. This speed-up effect is called “facilitatory interference”.

To sum up, Engelmann et al. (2015)’s pointed out, according to Lewis & Vasishth (2005), similarity-based interference effects causes inhibitory effects and is only present in target-match conditions. Misretrievals of the distractor decreases mean retrieval time causing facilitatory effects. It should be highlighted misretrievals happen in both target-match and target-mismatch conditions, while in target-match, the effect of similarity-based interference is assumed to be stronger.

After reviewing Lewis & Vasishth (2005) and comparing the materials and the results of previous research, the authors concluded experimental design might affect how prominent the distractor is. Distractor prominence can be affected by: (a) the linear order of target and distractor, distractors would be more active in memory when they are read more recently than targets; (b) the grammatical role of the distractor, distractors in the subject position evoke stronger interference than distractors in the object position; and (c) discourse saliency: when distractors are discursive topics, their interference effects are also stronger.

Besides distractor prominence, Engelmann et al. (2015) proposed that cue confusion should also be included in the extended model, since it is an independently motivated mechanism that can cause competition for activation even between different features. In other words, the authors defended similarity-based interference effects might occur even in target-mismatch conditions.

2.6.5 Weighting cues in memory

Van Dyke and McElree (2011) analyzed the role of similarity-based interference effects in memory by conducting a speed-accuracy tradeoff (SAT) study and an eye-tracking experiment testing subject-verb dependencies with intervening noun phrases (NPs). Their results pointed out the syntactic constraints at the retrieval site weighed more strongly than the semantic/pragmatic properties the potential distractors had in common with the target constituent. This way, the retrieval mechanism utilized a linear, weighed cue-combinatoric scheme, in which syntactic cues serve a gating function, so that only candidates in memory with matching syntactic properties are considered.

Dillon et al. (2013) also adopted the idea of relative weights to explain the differences found between reflexives and subject-verb dependencies. Their results indicated reflexives would follow a linear order of combinatoric cues, in which, syntactic cues would gate access to other cues such as the morphological ones; on the contrary, the cues involved in a subject-verb agreement dependency would be combined in a non-linear order so that syntactic cues would not block the influence of morphological cues shared by the target and the distractors. Thus, in reflexive dependencies, syntactic cues would weigh more than morphological cues, while in subject-verb dependencies syntactic and morphological cues would be similarly weighed.

2.6.6 Similarity-based interference in the encoding phase

While Lewis & Vasishth (2005) treated similarity-based interference effects during the maintenance of an item in memory, Oberauer & Kliegl (2006) and Lewandowsky

et al. (2008) was concerned about similarity-based interference effects during a prior phase - the encoding phase. According to Oberauer & Kliegl (2006), items are represented in working memory by sets of features that are activated together. If two items share the same feature, they compete with each other, and one the features will be lost in one of the sets (feature overwriting). This way, the item that lost a feature gets less distinguishable in memory, and the probability of its retrieval decreases. However, for Lewandowsky et al. (2008), a novel item in context might be assigned with a greater encoding weight; therefore, the greater the encoding weight, the easier it is to retrieve.

It should be emphasized that although interference effects may start in the encoding phase, it also affects the maintenance of an item in memory. Thus, it is pretty hard to distinguish encoding interference from maintenance interference.

2.6.7 C-command in content-addressable memory model

Kush et al. (2005) affirmed pronouns depend on the interpretation of antecedents in previous linguistic or non-linguistic contexts, and that accessing these antecedents in memory requires retrieval processes. The authors stated the relations between pronouns and their antecedents are subject of numerous constraints such as (a) the morphological constraints (e.g.: gender and number), which enforce feature-match relations between antecedents and pronouns; and (b) syntactic constraints (e.g.: c-command), which determine an antecedent's eligibility based on its relative structural position to a pronoun.

Kush et al. (2015) argued morphological constraints are item information, that is, they are drawn directly from a noun's lexical entry; therefore, they can be easily

encoded in memory as soon the noun is processed. However, the syntactic constraints of c-command are difficult to encode as inherent features of individual items because they refer to the relative position of X in relation to Y. Thus, encoding c-command would require look-ahead, or prediction of Y, which is not possible in the case of pronouns, as they are generally unpredictable. A solution for this problem would be encoding c-command relations for each item that is being introduced into the structure, requiring constant retroactive update of all c-commanding prior items. However, this would impose a significant computational burden on the parser.

Alcocer & Phillips (2012) were also concerned in understanding how c-commanding relations would be encoded in the content-addressable memory model. They proposed a series of ways in which c-command could be captured by a parallel access mechanism as described in Lewis & Vasishth (2005); however, none of them provided a solution that perfectly fits the goal of the model or existing empirical findings. Therefore, the authors concluded that serial mechanisms would be able to encode c-command, which does not exclude content-addressable model architecture. Alcocer & Phillips (2012) suggest a sequence of separate retrieval operations using individual node IDs as retrieval cues could be used to simulate a serial search of a tree. However, this process is neither able to explain empirical evidences of constant time access of retrieval in human parsing nor interference effects of non-c-commanding items.

Alternatively, Kush et al. (2015) reminded Discourse Prominence Theory encountered the same problem faced by c-command relations, that is, anaphora resolution holds iteratively searches through a list of NPs ranked according to their prominence. Therefore, a parallel dual retrieval strategy seemed adequate: one procedure that linearly traversed the c-command path to identify potential binders and

a second procedure that searched the linear string or a list of NPs ranked according to their prominence. However, not all antecedents need to c-command pronouns. Thus, it would be necessary two distinct mechanisms one for c-commanding antecedents and the other for non-c-commanding antecedents. As one can notice, this sounds quite difficult to happen.

Alternatively, Kush et al. (2015) proposed an item's status as a viable antecedent could be encoded as a feature, [± accessible], on a chunk in memory. The accessible feature could be used as a retrieval cue, that is, only those NPs that match the feature ACCESSIBLE should be retrieved as potential antecedents for a pronoun. This only would require updates whenever the parser shifts to a higher level of embedding from its previous position. The ACCESSIBLE cue is conceived as a simple precondition on antecedent-pronoun relations. It marks whether a phrase, could support some kind of anaphoric relation with a specified anaphora in the subsequent discourse. It should be mentioned that ACCESSIBILITY couldn't be used to enforce anti-locality constraints like Principle B, as well as it does not distinguish between c-commanding and non-c-commanding NPs.

ACCESSIBILITY can be a problem though for quantificational binding, in which antecedents must c-command the pronouns. Kush et al. (2015) explain that in this case, a post-retrieval stage would be needed in order to ascertain whether binding is possible. Therefore, the proposal presented by Kush et al. (2015) avoids encoding c-command constraints as a retrieval cue.

2.6.8 Summary

This section presented the differences between the multi-store memory models and the unitary-store memory model. In this dissertation, we will rely on the unitary-store memory model since it is the type of model that predicts a content-addressable memory approach, which is capable of explaining similarity-based interference effects in binding processing.

We reviewed two types of models concerning memory maintenance: the content-addressable memory model (ACT-R) proposed by Lewis & Vasishth (2005) and the working memory model by McElree and colleagues. On the one hand, the former model is a cue-based memory model with parallel retrieval, that is, speed remains constant irrespective of the number of competitors. However, the probability of successful retrieval decreases according to the number of competitors. Moreover, if misretrievals occur, reanalyses take time, which also influences retrieval speed. However, the latter model is based on the fact that spread activation would be divided between all items that match a certain cue. The item that receives the most spreading activation has the highest boost of activation, reaching the activation threshold first, facilitating retrieval. If there are competitor items that match the retrieval cues, the target receives less activation, and it is retrieved more slowly.

The extended content-addressable memory elaborated by Engelmann et al. (2015) defended that distractor prominence and cue confusion factors can better explain the diverging results reported in the literature. Thus this type of factors will also be taken into account in this dissertation.

Finally, with respect to the difficulty in turning c-command into a retrieval cue, the present work will assume the feature [+ accessible] to be the best way to represent binding relations.

2.7 Structural and morphological cues in binding processing

In this section, some previous research on coreferential processing with respect to Principles A and B will be reviewed[3]. The relationship between these structural constraints and agreement cues in the time-course of binding processing is very controversial in the literature. Therefore, previous research will be presented here under two subsections: works that showed some evidence of initial infallibility of structural constraints in binding processing; and works that found the opposite, that is, structurally unacceptable candidates can be initially considered as potential candidates if they feature-match the anaphoric expressions.

2.7.1 Evidences of initial infallibility of structural constraints in binding processing

Nicol & Swinney (1989) conducted a cross-modal priming experiment examining the reactivation of anaphoric antecedents. They found out that immediately after the anaphora only the structurally appropriate antecedent was reactivated, while the other referents were not significantly reactivated. The results for pronouns were similar to the results for anaphora. Thus, the authors concluded that the reactivation of prior referents is restricted by grammatical constraints. Nicol & Swinney (1989) explained that only when binding constraints do not constrain the list of potential antecedents to a single one; pragmatic and other sentence or discourse processing procedures would come into play, but only at a later point in processing.

[3] Principle C processing will not be addressed here since, unlike Principles A and B, it does not have any structural constraints operating on it.

Clifton et al. (1997) studied how antecedents of "her" and "him/his" are reactivated. They performed a phrase-by-phrase self-paced moving window experiment contrasting noun phrase (NP) and specifier (SPEC) usages. They also manipulated the morphological number of the subject in each sentence. The authors found faster reading times for the SPEC trials when the number of the subject agreed with the pronoun, which would make it an appropriate antecedent. However, when the subject and the pronouns mismatched in number, there was a slowdown on reading times, as the subject was made inappropriate. Importantly, number did not show any effects on NP trials. Thus Clifton and colleagues concluded that, at least initially, binding principles constrain parsing decisions, and that number would work as a filter to determine whether the accessed antecedents are appropriate.

Sturt (2003) was concerned about two questions: i) to what extent sentence processing is affected by ungrammatical antecedents; ii) to what extent do binding principles act like a filter on the final interpretation of a sentence. He conducted an eye-tracking study to investigate the influence of inaccessible antecedents in reflexive binding when they are put strongly into discourse focus. Stereotypical subjects were used in order not to expose participants to ungrammatical sentences. His results show that binding constraints were applied extremely early (at First Fixation and First Pass reading times). First Fixation and First Pass reading times were faster when the gender of the reflexive matched the stereotype of the accessible antecedent than it did not, but they did not differ reliably as a function of whether the inaccessible antecedent matched the reflexive. However, reliable influences of the inaccessible antecedent at late measures were found (Second Pass in the second area after the reflexive). There were longer Second Pass times when the inaccessible antecedent mismatched the reflexive than when it did not. The author concluded that antecedents

that were not initially considered by the binding principles could affect processing at a later stage. In other words, binding constraints are applied at an extremely early stage, but they do not act as filters. Sturt (2003) also conducted a follow-up study, a sentence-by-sentence self-paced reading experiment with a comprehension question to check the interpretation of the anaphora referent. It seems that Principle A did not act as an absolute filter on the final interpretation of the sentence either. Sturt (2003) defends the idea that binding principles act like a defeasible filter, as they can be violated at a later stage when there is a highly focused inacceptable antecedent involved.

Leitão et al. (2008) investigated the relationship between Principle B and phi-features (gender, number, and animacy) in coreference processing in Brazilian Portuguese in two self-paced reading experiments. In the first experiment, there were structurally unacceptable antecedents in the sentences, and the results showed that the pronoun+1 region (adverb regions) had longer reading times due to the fact that the structurally unacceptable antecedent in the sentence feature-matched the pronoun. However, in the second experiment, there was a structurally unacceptable candidate available in a preamble. Unlike the first experiment, the results of the second experiment did not show any differences among the conditions, although the reading times at the pronoun region were faster when compared to the first experiment. The authors suggested that when there are no structurally acceptable antecedent candidates available, as in the first experiment, candidates that feature-match the pronouns could be considered as potential antecedents even if they violate Principle B. However, when there is a structurally acceptable antecedent available, as in the second experiment, the search of an antecedent ends faster and the structurally unacceptable candidates are not taken into account.

In an *Event Relative Potentials* experiment (ERPs), Xiang et al. (2009) studied intrusion effects of structurally unacceptable noun phrases that matched the reflexive. The authors found a P600-like component for both intrusive and incongruent conditions. However, there were no differences between the intrusive and incongruent sentences, while both were significantly different from the congruent. It is important that they found a marginal late intrusion effect only at 800-1000ms, which matches the late effects of inaccessible antecedents in Sturt (2003). The authors concluded that there is no initial intrusion effect for reflexive binding.

Oliveira et al. (2012) conducted a self-paced reading experiment to determine whether Principle A influences reflexive resolution in Brazilian Portuguese. They found that the grammatical conditions, in which the structurally acceptable antecedent agrees in gender with the reflexives, had faster reading times at the reflexive region when compared to ungrammatical conditions. It should be noted that the structurally unacceptable antecedents were not taken into account in any condition, which suggests that Principle A works as a filter, blocking the candidates that violate it.

Dillon et al. (2013) conducted eye-tracking experiments and grammaticality judgment tests with the purpose of investigating the impact of structurally illicit nouns phrases on the computation of reflexive binding. The off-line results of the grammaticality judgment tests indicated sentences with number mismatch between reflexives and structurally licit antecedents were judged as ungrammatical by the participants despite of the presence of structurally illicit nouns phrases number matching the reflexives. Similarly the on-line results of the eye-tracking experiments did not find any facilitatory intrusion effects caused by structurally illicit noun phrases, that is, the presence of number matching structurally illicit antecedents could not facilitate retrievals even when structurally licit antecedents mismatched the

reflexives. The authors concluded that the feature content of a structurally illicit noun phrase could not affect reflexive processing. They concluded that the mechanism used by reflexives does not use morphological cues; it rather uses binding constraints syntactic information to guide antecedent retrieval.

Chow et al. (2014) were concerned about which kinds of constraints initially restrict antecedent retrieval, and which have later effects, working as filters. In their first self-paced moving window experiment they manipulated the gender match between the pronoun "him" and the structurally acceptable main clause subject and the structurally unacceptable embedded clause subject. Relative clauses could also modify the nouns in order to increase the distance between the pronoun and the antecedent. The structurally unacceptable antecedents could be either a common noun or a proper name. As the mismatch conditions had longer reading times, it seems that comprehenders are immediately sensitive to the structural constraints on pronoun interpretation regardless of the similarity between the candidate antecedents and linear distance. They found robust effects of grammaticality, but no interference effects of any kind. It should be mentioned that when the linear distance between the pronoun and the structurally acceptable antecedent was long in the modified common noun condition, they found a late ungrammatical match effect, that is, when no grammatical antecedent was available, the presence of a feature-matching structurally unacceptable antecedent led to longer reading times. The authors explain that it may have been caused by the fact that the memory representation of the structurally acceptable antecedent was decayed due to the long distance. In their second experiment, Chow et al. (2014) tried to replicate the results found on Badecker & Straub (2002) [which will be discussed in the next subsection] by using identical materials and procedures. However, Chow et al. (2014) failed and only replicated the

results of their first experiment. They also conducted 3 other experiments, but no effects were found. The authors defended the *Simultaneous Constraints hypothesis* since it appeared that both agreement features like gender and the structural constraints of binding immediately restricted the set of candidate antecedents during the initial retrieval process.

Cunnings et al. (2015) investigated how syntactic constraints and gender agreement interact to guide memory retrieval during the resolution of subject pronouns. The authors' primary goal was to investigate whether c-command constraint restricts antecedent retrieval. They conducted a series of eye-tracking experiments on pronoun resolution with quantified and non-quantified antecedents. Their results confirmed that non c-commanding antecedents were not ignored due to their lower discourse salience. They found out both c-commanding and non c-commanding antecedents can be readily retrieved, but only when they are structurally acceptable antecedents. Furthermore, they showed structurally acceptable candidate are always preferably retrieved than structurally unacceptable candidate, even the structurally acceptable candidate mismatched the pronoun in gender. This is evidence to the fact that c-command restricts antecedent retrieval during anaphora resolution. The only case in which structurally unacceptable candidate seemed to be taken into account was in regression path times for the final region: reading times were significantly longer when the structurally unacceptable QP candidate matched the gender of the pronouns, but only when the structurally acceptable antecedent mismatched the gender of the pronoun.

2.7.2 Evidence of initial fallibility of the structural constraints in binding processing

Badecker & Straub (2002) studied the processing of reflexive and pronoun binding in a series of self-paced reading experiments. According to the authors, coreference processing is influenced by: morphological and syntactic properties of the dependent expression and the antecedents; structural parallelism; causal semantics; prominence and salience of the local discourse entities; and the world knowledge shared about the discourse entities involved. Among these factors, the authors' study was focused on morphosyntactic features and local focus of attention. In one of their experiments, they investigated whether the content of structurally inaccessible NPs would influence pronoun processing.

(68) Sample of the materials in Badecker & Straub (2002)

a. multiple match: John thought that Bill owed **him** another chance to solve the problem.

b. accessible match: John thought that Beth owed **him** another chance to solve the problem.

c. inaccessible match: Jane thought that Bill owed **him** another chance to solve the problem.

d. no-match: John thought that Beth owed **him** another chance to solve the problem.

They observed longer reading times in the no-match condition than in the accessible match condition. The results also show faster reading times when there was

a structurally accessible antecedent than when there was an inaccessible antecedent. There was no difference between the multiple match and the accessible-match conditions. The authors concluded that gender was automatically used to identify the referent of a pronoun, and that the structurally accessible antecedents were also rapidly accessed. In contrast, inaccessible candidates were not blocked for an initial candidate set, as they influenced the evaluation process as soon as the pronoun was encountered.

Badecker & Straub (2002) also investigated whether number features could shape the initial candidate set. In another experiment, they studied the influence of grammatical number in reciprocal anaphors like "each other", which are also governed by Principle A, as can be seen in (69):

(69) Sample of the materials in Badecker & Straub (2002)

a. multiple match: The attorney thought that the judges were telling **each other** which defendants has appeared as witnesses before.

b. single-match: The attorneys thought that the judges were telling **each other** which defendants has appeared as witnesses before.

The results indicate longer reading times in the multiple-match than in the single match, but only 3-4 words after the anaphor. The authors suggested that morphological number contributes to identifying the initial set of antecedent candidates. The multiple-match effect was attenuated in this case, because, according to the authors, common nouns may not be as effective as proper names in establishing discourse entities.

Badecker and Straub (2002) concluded that binding-theory principles do not function as initial filters as reading times were longer when the grammatically inaccessible NPs agreed in gender (and number) with the pronoun or anaphor. The authors supported the *interactive-parallel-constraint model*: the initial candidate set is composed of the focused discourse entities that are compatible with the lexical properties of the referentially dependent expression, while the grammatical constraints on interpretation operate quickly and effectively in the process of selecting from among these options.

Kennison (2003) investigated how comprehenders use structural information during coreference resolution of the pronouns "her", "him", and "his". In a self-paced moving window experiment, Kennison (2003) examined the processing of "her" in object position, functioning as either a SPEC as in (70) or an NP as in (71).

(70) Sample of the materials in Kennison (2003)

SPEC conditions:

a. Susan watched **her** classmate during the open rehearsals of the school play.

b. Carl watched **her** classmate during the open rehearsals of the school play.

c. They watched **her** classmate during the open rehearsals of the school play.

NP conditions:

a. Susan watched **her** during the open rehearsals of the school play.

b. Carl watched **her** during the open rehearsals of the school play.

c. They watched **her** during the open rehearsals of the school play.

She found that the type of subject influenced coreference processing in both conditions, including in NP conditions, which is inconsistent with Nicol & Swinney (1989) and Clifton et al. (1997). In SPEC conditions, reading times were longer when the subject was a male name, while in NP conditions reading times were longer when the subject was female. And the shortest times were for the conditions with "they". In other words, when coreference could be achieved, there were longer reading times for NP conditions than for SPEC, as SPEC conditions were easy to process. However, when coreference could not be achieved, there was no difference immediately after the pronoun. But, later, when gender and number information was accessed, coreference was impeded in SPEC sentences as reading times were longer for the SPEC than the NP condition later on in the sentence. Kennison (2003) also replicated the results of "her" with "his".

Kennison (2003)`s findings contradict Nicol & Swinney (1989) and Clifton et al. (1997) as structurally unavailable antecedents were considered as potential subjects since the type of subject influenced reading times. Her findings also contradict Badecker & Straub (2002), as number features appeared to help compose the initial candidate set, while gender mismatch only influenced processing at a later phase. It seemed that the antecedent search ended more quickly when the unavailable candidate differed in number with the pronoun whereas the antecedent search was longer when the subject of the sentence in NP matched the pronoun in gender.

In another experiment, Kennison (2003) aimed to determine whether subject type would influence processing when the discourse context contained an available antecedent for the pronoun as in (71).

(71) Sample of the materials in Kennison (2003)

Billy complained about having a stomachache.

a. Laura watched **him** closely throughout the day.

b. Michael watched **him** closely throughout the day.

c. They watched **him** closely throughout the day.

The results suggested that when a single highly salient and structurally available antecedent was in discourse context, structurally unavailable antecedents did not influence coreference, which means that when there is a good fit between the antecedent and the pronoun, the process of searching for an antecedent terminates. It appeared that, on the contrary, when no antecedent is available or when there is not a strong fit between the structurally available antecedent and the pronoun, the process of searching for an antecedent continues, and structurally unavailable antecedents can be considered.

Cunnings & Felser (2013) aimed at distinguishing effects arising because of structural constraints of reflexive binding from those arising because of working memory limitations. The authors believed the contradictory results in the literature might have to do with the differences of working memory span that exist among participants. They conducted two eye-tracking experiments manipulating the linear distance between the antecedents and the reflexives. The results found by the authors indicated a clear evidence of Principle A guiding the earliest stages of processing in either short or long distance conditions. The effect of Principle A was even earlier for participants with low working memory span than for participants with high working memory span. Effects of structurally unacceptable candidates were only encountered at later stages of processing in short distance conditions. However, when there was a long distance between antecedents and reflexives, low working memory span

participants were influenced by both structurally acceptable and unacceptable antecedent candidates during the earliest stages of processing. Cunnings & Felser (2013) explained when low span readers encountered the reflexives, they were early guided by Principle A, but they had difficulty in inhibiting the activation of the structurally unacceptable competitor antecedent. This difficulty might have been caused by the fact the structurally unacceptable candidate was the subject of a relative clause, which sounds more discourse prominent than other studies reported in the literature. Moreover, the structurally unacceptable candidate in their materials was actually a pronoun that referred back to the character mentioned in the introductory sentence, which might have made this candidate even more prominent in discourse.

Parker (2014) studied how the parser targets specific information in memory, and how that information is extracted to elaborate the sentence representation. The author studied attraction effects in anaphora resolution manipulating gender, number, and animacy. The results for 1-feature mismatch only showed a late slow down in reading times for ungrammatical sentences, and no attraction effects were found. However, for 2-feature mismatch conditions, early and late reading times were facilitated for ungrammatical sentences with attractors when compared to ungrammatical sentences without attractors. Parker (2014) explains that attraction effects are likely to be a consequence of quantitative similarity. Qualitative factors are also important since structural cues are weighted more strongly in retrieval than morphological cues.

Patil et al. (2016) thought that reflexive binding might be a very informative phenomenon in understanding the role that grammatical and non-grammatical constraints play in memory. The structural constraints of reflexive binding are relatively clear, and this construction admits manipulations of agreement, distance,

and distracting antecedent candidates. They created a model running 1000 simulations of each condition of Sturt (2003)'s conditions. On the one hand, just like Sturt (2003), they found that: retrieval errors on mismatch conditions were higher than in match conditions (mismatch effect), the retrieval errors for both interference conditions, mismatch and match, were higher than for the other 2 conditions (match interference effect), and the retrieval times for both mismatch conditions are longer than the other two match conditions (mismatch effect). On the other hand, they also found results that were not consistent with Sturt (2003): retrieval times for the match interference condition were shorter than for the match condition and shorter than for the mismatch conditions (mismatch interference effect). Patil et al. (2016) suggested that the inacceptable candidates in Sturt (2003) could not be good attractors as semantic matching cues are not able to cause attraction if no grammatical cue is involved. In addition, since they were less recently created in representation, they could not have enough strength in memory to be retrieved due to decay factors.

Patil et al. (2016) also conducted an eye-tracking experiment. To increase the strength of the inaccessible subject, they used an object pronoun within a relative clause where the inaccessible antecedents were the subjects of the clause. Patil et al. (2016) found a significant main effect of interference in First Pass and in First Pass Regression Probability. Patil et al. (2016) concluded that non-structural cues such as gender are crucial for antecedent retrieval so that gender agreement features must be included in the set of retrieval cues. Moreover, it seems their results are inconsistent with strict syntactic constraints on antecedent retrieval, and it seems reflexive binding is not infallible.

2.7.3 Summary

As already mentioned, the literature on the relationship between binding structural cues and agreement structural cues is very contracting. Some studies claim memory cannot be initially influenced by distractors [Nicol & Swinney, 1989; Clifton et al. 1997; Sturt, 2003; Xiang et al., 2009; Dillon et al., 2013; Chow et al., 2014; Cunnings et al., (2015), among others]. While others defend the opposite, that is, the early fallibility of binding structural cues [Badecker & Straub, 2002; Kennison, 2003; Cunnings & Felser, 2013; Parker, 2014; Patil et al., 2016, among others].

However, it should be mentioned Engelmann et al. (2015) proposed an extension to the content-addressable memory model that seemed to accommodate all those diverging results: distractor prominence and cue confusion. The authors also tried to explain all the conflicting results regarding slower and faster retrievals. This new model will be addressed in the discussion of the experiments of this dissertation.

2.8 Chapter digest

This chapter focused on reviewing different theories and experimental data that investigated binding and memory. Binding is a central topic in linguistics for both theoretical linguistics and psycholinguistics since it allows an understanding of how structural, syntactic, meaning, and discursive factors work together in order to establish binding / coreference, which is present in every language and is the main object of study of this dissertation.

Among the three binding theories presented in this chapter, namely the Standard Binding Theory (Chomsky, 1981, 1986, 1993) and the two Predicate-based Binding Theories by Pollard & Sag (1992, 1994) and Reinhart & Reuland (1993), the

Standard Binding Theory, despite its flaws, was chosen as the theoretical basis in this dissertation. Thus the concept of structural binding cues that will be often referred to is the one proposed by Chomsky (1981, 1986, 1993). Principle B will be especially important for this dissertation: a pronoun must be free in its governing category, often called local clause.

From the discursive binding theories, the notions of prominence, focus of attention, topichood, and accessibility [Gordon & Hendrick, 1998; Grosz et al., 1995; Arnold et al., 2000; Foraker & McElree, 2007; Greene et al., 1992; Rigalleau et al., 2004; Kehler, 2007] will be essential to understand the materials used in the present work. Antecedent candidates that occupy the subject position are more prominent than other candidates (Gordon & Hendrick, 1998). In addition, they are generally considered to be the discursive topic too (Grosz et al., 1995). However, local antecedent candidates can be the focus of attention (Foraker & McElree, 2007) and considered to be more accessible (Arnold et al., 2000) due to recency factors (McElree, 1996). Thus the materials used in experiments of this dissertation will combine locality and subjecthood bias: structurally unacceptable antecedent candidates will be subjects of an embedded relative clause followed by an object pronoun located in the same clause.

The present work will rely on the unitary-store memory model (Anderson, 1983; Cowan, 2001; McElree, 2001) and the ACT-R content addressable memory model (Lewis & Vasishth, 2005; Lewis et al. 2006). The unitary-store memory model posits that only one type of memory exists: the long-term memory. According to this model, memory is bi-partite between active and passive memory. Items from passive memory constantly move to active memory (focus of attention). And it is the working memory the responsible for this efficient interchange. The ACT-R content

addressable memory model is conceived under the unitary-store memory model. According to this approach, items are parallel retrieved in memory according to their contents. Therefore, if more than one item shares the same contents with the target, it causes similarity-based interference effects. Consequently, retrieval times are slower due to the competition between the items, and misretrievals can occur as the latency of activation decreases with more than one item being activated. In the case of misretrievals, faster retrieval times are predicted.

From the previous research on interference effects reviewed in this chapter, this dissertation is in line with those studies that encountered early fallibity of binding structural constraints, that is, when structurally unacceptable antecedent candidates initially influenced binding resolution.

3. Gender cues in word processing

Gender is the most puzzling of the grammatical categories. It is a topic which interests non-linguists as well as linguists and it becomes more fascinating the more it is investigated.
(Corbett, 1991, p.1)

3.1 Outline

This chapter assesses the concept of gender in Linguistics as it provides ways to understand the lexicon structure and the role of morphology in language. More importantly it reveals interesting aspects of the human mind and how it refers to sexed beings and inanimate referents in the world.

In the first section "How words are formed", basic concepts and terminologies on morphology will be presented. The importance of roots and affixes will be explored as well as the complicated relation they have with each other. In the second section "Theoretical approaches on morphology representation", generative proposals on the place of morphology in grammar will be briefly described. In the third section "Morphological processing", some pioneer studies on morphological processing will be presented. They provide empirical evidences for theory and some of them guide experimental research until nowadays. From the section "Genders in the languages of the world" on, gender will be in the focus of discussion in this chapter. In this fourth section, Corbett's remarks regarding the gender systems across the languages of the world will reveal particular nuances of gender that will be addressed in the present study. In the fifth section "Gender in Portuguese", two important linguists who are separated by some decades across time will discuss how gender system in structured

in Portuguese. The sixth section "Unmarked and default forms in gender" will examine the strategies languages with masculine and feminine systems use to have neuter gender semantics. The seventh section "Gender agreement in Portuguese" will illustrate why Portuguese is known as a language with a rich morphology system. The eighth section "Previous research on gender processing" will review some psycholinguistic findings on gender processing in Portuguese as well as in languages that seem closer to Portuguese as Italian and Spanish. Finally, the last section will be focused on discussing experimental evidence on different types of gender: grammatical, semantic, definitional, and stereotypical.

3.2 Introduction

From the quote in the beginning of this chapter, one can realize Greville Corbett (1991) begins his book entitled *Gender* asserting not only the importance of gender for some languages of the world, but also describing some practical benefits that studies on gender can bring in short and long terms. According to him, gender may be a problem for speakers who are learning a foreign language; therefore, studies on this grammatical category can contribute to second-language learning. Studies on gender can also reveal how linguistic information is stored in the brain and how natural language processing takes place in eliminating, for example, local ambiguities.

Corbett (1991) defines the word "gender" using its etymology origin, *genus*; via Old French, *gendre*, meaning "kind" or "sort". He explains gender is used for a group of nouns, but also for the whole category. This is the reason why a particular language is said to have 3 genders (masculine, feminine and neuter), or that a language has the category of gender.

Gender sounds very puzzling if one considers questions such as “How are words allocated to different genders?” or “How do native speakers of a language know the gender of a particular word?” Corbett (1991) states it would be impossible for speakers to remember the genders of all words, not only because memory is limited, but also because more errors would be expected. But what happens is that native speakers make few or no mistakes in gender. In addition, if it were a matter of memory, “How would borrowings from other languages or even invented nouns receive a gender so quickly with that degree of consistency?” To answer those questions, it is crucial to understand how words are structured and represented in our minds/brains. The next section will address the object of study of morphology, followed by an overview of the main theoretical approaches and how linguists manage to explain lexical knowledge.

3.3 How words are formed

According to Dupoux (1998), in spoken language, the word is the smallest unit of communication; however, it is not the smallest verbal gesture capable of expressing meaning. A single “s” can make all the difference at the end of the word “fact”, for example. That “s” expresses plural, which means there are more than one fact. Therefore, words can be broken down in even finer units – the morphemes. The study of which kinds of morphemes combine with other kinds of morphemes, and how this affects meaning of the words is the object of study of morphology.

There are different kinds of morphemes: roots, which are the core of the word, the smallest unit in common with other derived words; and affixes, elements with already defined syntactic-semantics properties that constrain the meaning and use of

the new-formed word. In addition, stems are composed by roots and affixes, and they work as bases for other affixes so that new words are created. In the example above, "fact" is the root and "s" is the affix. Affixes that come after the root are called suffixes, like the "-s" in the previous example. Affixes that come before the root are called prefixes. And affixes that are inserted into the root are called infixes. Languages differ in terms of where they put their affixes. In Indo-European languages, the most common affixes are the suffixes.

Dupoux (1998) argued affixes are not the only way to modify the word meaning, stems can have one of the vowels changed, as for example, in the past tense of the some verbs in English like "run – ran", "speak – spoke". This is a remnant from the time past tense was not formed by adding *–ed* to the verbs. In Semitic languages, like Hebrew, this kind of modification is very common though.

Dupoux (1998) also pointed out there are different kinds of affixes. Inflectional affixes do not change the meaning of the word, but convey additional information; while derivational affixes change the meaning and are used to derive new words. Examples of inflectional suffixes are "-s" in "fact" and "-ing" in "governing". An example of a derivational suffix is "-able" in "governable", while "un-" in "ungovernable" is an example of a derivational prefix. All of these examples show words with related meaning, although each of them means something different. However, several words that seem to share a root are unrelated in meaning, as for instance, "depress" and "express" or "casual" and "casualty". In addition, some words cannot have their meanings deduced from taking the affixes away, as "submit" cannot be deduced from "mit".

As one can see, words are complicated. There are a huge number of variables involved in the way words are formed. And some questions remain: "How can

speakers of a language know all those thousands of words?" and "How can we recognize words so quickly when listening to someone or when reading?". For decades, linguists have been trying to explain how morphology is represented in grammar. In the next section, some of the most important approaches on morphology in the generative grammar theory will be discussed.

3.4 Theoretical approaches on morphological representation

According to the Lexicalist Hypothesis (Chomsky, 1970) the lexicon is the repository of all the idiosyncratic properties of the lexical items, including the phonological forms, the syntactic category specifications, the semantic characteristics, and the syntactic properties related to the argumental structure of the items. According to Chomsky, all the information regarding the internal word structure is stored in the lexicon, and syntax cannot operate on them. The syntactic computational system only takes from the lexicon the primary symbols, which are atomic structures, containing the sequence for the syntagmatic structure to be formed in the tree. In a nutshell, Chomsky defended the idea syntax is independent from morphology.

Anderson (1982) claimed the decline of the study of morphology was caused by the development of the generative grammar of Noam Chomsky. He argued Chomsky's proposal obviously could not explain the facts of natural language.

> "Elements of morphological structure and of sentence structure are not totally unrelated; some of the properties of individual words are essentially dependent on their position in larger structures, and some principles operating over domains larger than a single word must be able to refer to properties of the words themselves (ANDERSON, 1982, p. 573)."

Anderson (1982) argued some words have morphological properties that directly depend on their position in the sentences as well as on other words within a sentence. Thus if adjectives agree with nouns, the syntactic rule of agreement must have access to the features of gender, number, etc on which agreement is based. In other words, word-internal elements are strongly interconnected with syntax processes.

In addition, the author highlighted the importance of differentiating what is inflectional from what is derivational. For Anderson (1982), morphology can be divided into inflection and derivation. The former is inextricably integrated with syntax, while the latter is confined to the lexicon, independent of syntax. Following Anderson's ideas, Pearlmutter (1988) formulated the Split-Morphology Hypothesis. In this model, derivation should be handled by lexical rules, whereas inflection should be handled by syntactic rules.

On the one hand, Halle & Marantz (1993, 1994) agreed neither with the Lexicalist Hypothesis nor to the Split-Morphology Hypothesis, but they used some ideas of those two models and postulated a new theory, which is called Distributed Morphology, also known as a non-lexicalist or constructionist hypothesis. For Distributed Morphology, inflection and derivation should be treated similarly and structured by syntactic processing. This way, syntax operates "all the way down" and sentences are generated by the same mechanisms presented in the Minimalist Program (starting with Chomsky, 1993): "merge" and "move". In this view, syntax would operate not only on sentences, but also within the words, merging roots and morphemes. In Distributed Morphology, the roots and morphosyntactic features stored in List 1 (poor lexicon) generate an abstract syntactic structure, which will require phonological interpretation from List 2 (vocabulary). And after that, the word

units will eventually be associated with the meanings stored in List 3, also called Encyclopedia.

On the other hand, Di Sciullo & Williams (1987) elaborated a model in which words are syntactic atoms, that is, syntax cannot analyze the internal structure of words, but it can analyze their features. Words only verify whether their features match the features of the syntactic structures. In this view, syntax and morphology are parallel, but independent. This way, morphology is neither confined in the lexicon nor a syntax parasite, but it is an independent component of language.

It is preferable to think of morphology and syntax as two independent systems, following the same spirit of Di Sciullo & Williams (1987). It is hard to image all morphological diversity that exists in the languages of the world would be constrained by either the lexicon or syntax. Morphology is too much valuable for that. It is vital for the languages and it deserves a good space in the linguistic theory.

3.5 Morphological processing

Psycholinguists have debated whether word recognition would be based on the features of the whole-forms or on individual features of components of the words – the morphemes. By analyzing a word like "unlucky", Taft & Forster (1975) affirmed this word would probably be stored in the lexicon in conjunction with its base word "luck" (along with "lucky", "luckily", "luckless", etc.). Moreover, they claimed there would be no separate lexical entry for "unlucky" in the lexicon, since this word is formed from the entry "luck" and the affixes "un" and "y". This way, the word "unlucky" would require a prior morphological analysis, that is, the affixes would be stripped off so that the base form "luck" would be accessed. This is known as the

Affix Stripping Hypothesis. In contrast, Butterworth (1983) defends the Full Listing Hypothesis, which posits every word-form is explicitly listed in the mental lexicon, including inflectional or derived complex words. In this model, the word is the basic lexical unit rather than the morphemes. Consequently there would not be any difference in how complex and simple words are stored.

The model formulated by Marslen-Wilson and colleagues can dialogue with the Full Listing Hypothesis. During the 1980s, Marslen-Wilson conducted a series of studies in order to explain how spoken words are converted into meaning. According to the Cohort Model, listeners can recognize the words before their ending. Marslen-Wilson (1987) argued at 100-150ms after the word onset, all words that match that initial sequence are activated in parallel. High-frequent words would be stronger candidates than low-frequent words since their level of activation would be higher. As the structural and interpretative context is continued monitored, there would be a sequential reduction of the initial candidate set until only one candidate is left. At this point, known as the *uniqueness point*, the correct candidate would be recognized and incorporated into representation. Modern versions of the Cohort Model predict words can be recognized even before the uniqueness point.

Caramazza et al. (1988) postulated The Augmented Addressed Morphology Model. This is a quite interesting name for a model. The reason for using the terms "addressed morphology" is because known words (already experienced words) may be fully accessed in the lexicon, without decomposition. They used the term "augmented" because the model included in the analysis novel words, that is, words that have not been previously experienced, which, according to the authors, are morphologically accessed though. This model defends lexical processing takes a dual-

route mechanism, that is, there are two distinct processes that may vary depending on the nature of the lexical items.

A dual-route mechanism model was also defended by Pinker (1991) and Ullman et al. (1997). According to him, two capacities are the source of the expressive power of human language: the mental lexicon and the mental grammar. The first contains thousands of words, memorized with arbitrary sound-meaning pairing. The second contains generative rules that combine words into an infinite number of phrases and sentences.

Pinker (1991) concluded regular verbs (walk-walked) are computed by suffixation rule, whereas irregular verbs (run-ran) are retrieved from an associative memory. Furthermore, with the results of a study on aphasic patients, Ullman et al. (1997) confirmed his previous findings: irregular words are accessed by the mental lexicon, while regular and novel words are accessed by the mental grammar.

Allen et al. (2003) found out evidences in favor of a dual-route mechanism in morphology processing by carrying experiments on ERPs[4]. As Pinker (1991), they studied the regular and irregular past tenses in English, but they controlled for the frequency of the forms. They authors found out the inflectional features of regular forms were only accessed after morphological parsing. Regular low-frequent forms elicited a more negative N400[5] than regular high-frequent forms. Moreover, it was detected a more positive P600[6] in the ungrammatical condition than in grammatical condition. With respect to irregular forms, Allen et al. (2003) did not found any N400

[4] ERPs stands for Event-Relative Brain Potentials and correspond to small voltage changes in the electrical activity of the brain, recorded from the scalp, consistently triggered by an external stimulus or a cognitive event.

[5] N400 is a negative signal that appears at approximately 300-400ms after the stimuli is presented. It is generally associated to lexical access.

[6] P600 is a positive signal that appears at approximately 500-600ms after the stimuli is presented. It is generally associated to syntactic violations and reanalysis.

effects, but they detected the onset of P600 appeared earlier and with a larger amplitude for the high-frequent forms than for the low-frequent ones for the ungrammatical condition. They concluded irregular forms provide direct access to inflectional features without morphological parsing. In addition, the authors argued decomposition seems more processing demanding than direct access, but it is more flexible once it allows lexical content information to be processed independently from inflectional parsing.

Ullman (2001) argued in favor of a connectionist model. According to him, there is one-route mechanism responsible for language processing – associative memory, which learns, represents and computes all lexical and grammatical knowledge. The associative memory is domain general, that is, it is not exclusively used for language. In this theory, there is no difference between non-compositional and compositional forms, as well as there are no mental rules and no system to process those rules. Learning this model occurs by adjusting weights on connections on the basis of statistical contingencies in the environment (Rumelhart & McClelland, 1986; Seindenberg & McClelland, 1989, among others).

In this section the role of morphology in word processing was discussed and the main word processing models were briefly reviewed. The conclusion is the literature is very contradicting and psycholinguists have not reached a final conclusion regarding word retrieval/access. This way, more evidence is needed, especially in languages with rich visible morphology, as the Romance languages.

Maia et al. (2007) was interested in examining how complex words are stored and accessed in Portuguese. Their goal was to investigate whether there is morphological decomposition before the words are accessed. In Experiment 1, the authors conducted a color decision task with *stroop effects*. In this experiment,

isolated words were presented with some letters in different colors. The authors expected that the color of the letters would be easily recognized by the participants if they were part of a morpheme, which would be colored in the same color. The authors also controlled for the type of morpheme, which could be morphemes merged into words (semantic transparency between the derived word and the base, for example, *malinha* (small suitcase), pseudo-morphemes (only orthographic coincidence with morphemes, for example, *espinha* (spine), or morphemes merged into roots (arbitrary meaning, no decomposition processed would be involved, for example, *caninha* (*cachaça*, which is an alcoholic beverage). The results indicated the participants correctly recognized the letters color when morphemes were merged into words or into roots. This is evidence readers would perform morphological parsing in cases in which morphemes were either transparent or opaque in relation to the base. In the cases of pseudo-morphemes, morphological parsing would not be activated, as there is no morpheme to be segmented. In Experiment 2, the authors conducted an eye-tracking experiment without *stroop effects* with the purpose of verifying whether words with morphemes would be more costly processed than words with pseudo-morphemes. The authors found out morphological parsing effects only for the condition in which morphemes were merged into words, which indicated that pseudo-morphemes and morphemes merged into roots were accessed through *Full Listing*. The authors concluded lexical access in Portuguese takes both top-down (*Full Listing*, Butterworth, 1983) and bottom-up (morphological decomposition, *Affix-Stripping*, Taft & Forster, 1975) processes, which is evidence in favor of a dual-route mechanism.

Maia & Ribeiro (2015) conducted an eye-tracking experiment with a lexical decision task comparing mono-morphemic and multi-morphemic words and pseudo-

words in Portuguese. The authors found out when the internal word structure can be segmented into morphemes, bottom-up processing (morphological parsing) is preferable rather than top-down (*Full Listing*) processing. In addition, multi-morphemic words and pseudo-words, as for example, jornaleiro (newsagent) and *norbalense* respectively, were more costly processed than mono-morphemic words and pseudo-words, as for example, *jabuticaba* (Brazilian grape) and *liboramima* respectively. Moreover, multi-morphemic words were more easily recognized than mono-morphemic words, and multi-morphemic pseudo-words were more accepted than mono-morphemic pseudo-words.

Garcia (2009) was concerned in examining the role of morphology in word recognition and in lexical organization. She conducted a priming experiment with lexical decision task with the purpose of verifying whether during an initial and automatic activation of a word, its morphemes would be also activated. Aiming at dissociating morphology from semantics and phonology, the author tested 4 priming conditions: (1) prime and target are morphologically related, as for instance in pairs like *fila* (line) – *fileira* (line); (2) prime and target are only semantically related, as in pairs like *ordem* (order) – *fileira* (line); (3) prime and target are only phonologically related, as in pairs like *filé* (fillet) and fileira (line); and (4) prime and target are not related, as in pairs like *mato* (grass) – fileira (line). As she expected, the fastest pairs to be recognized were in condition 1, in which the prime and the target shared the same root. Garcia (2009) concluded that the mental lexicon is morphologically organized.

One must keep in mind once gender is also morphologically expressed, studies on gender might contribute to the question of word processing. Next, an overview on how gender differs among the languages of the world will be presented.

3.6 Gender in the languages of the world

Linguists have been trying to understand how native speakers are capable of assigning gender to words. Understanding assigning can reveal how the lexicon is structured. For Corbett (1991), gender assigning depends on two basic types of information about the nouns: meaning (semantics) and form. The latter could be originated from morphology (word-structure - derivation, inflection) or from phonology (sound-structure). He highlights the idea gender systems are always semantic motivated though and languages can combine morphological and phonological factors at different ways allowing sets of exceptions.

Corbett (1991) classifies the languages of the world into 2 major systems: semantic and formal. In the first system, the meaning of a noun determines its gender, and consequently, given the gender of a noun, one could infer something about its meaning. Gender in those languages is assigned based on semantic criteria only. In the second system, rules depend on their form rather than their meaning. Those rules can be morphological or phonological. Thus formal systems can be divided into morphological systems and phonological systems. Although, there is not a clear distinction whether a rule is morphological or phonological, Corbett (1991) explains morphological rules are more complex as they need information from more than one form, while phonological rules refer to a single form of a noun. An example of a morphological rule in Russian is "nouns of declension type I are masculine", while an example of phonological rule in Portuguese would be "nouns ending in *–a* are feminine".

On the one hand, in languages with semantic system, gender is only encoded for elements that have biological sex. On the other hand, in languages with a semantic system, such as the Romance languages, all nouns have gender. Vigliocco & Franck

(1999) explains nouns that refer to animate entities generally have a transparent relation between gender and sex. This type of gender is called conceptual gender (also called semantic gender or natural gender). However, for many inanimate nouns (objects or abstract entities), animal names, or for a few nouns referring to humans, gender does not have any relation to the sex of the referent. And this type of gender is called grammatical gender.

Moreover, Cobertt stated when the gender of a noun is visibly evident from its form, it has overt gender; however, when its form does not visibly show the gender of a noun, it has covert gender. Languages with clear formal assignment systems have overt gender. It should be noticed overt and covert gender are extreme poles in a continuum. An overt gender language would have ideally, a marker for gender on every noun. Portuguese, for example, is mostly overt.

Corbett (1991) also refers to some problematic nouns in the languages of the world: the hybrids, the double- and multiple-gender nouns, and the epicenes. Hybrids are a result of a conflict between different assigning rules. They do not belong to a single gender. An example would the boat names in English. Although, boats may have a masculine name in English, they would agree with feminine pronouns like "she". It is a conflict between formal and semantic criteria. Another example in German would be *mädchen* (girl), which is semantically feminine, but it is morphologically neuter because of the diminutive *–chen*. Thus *mädchen* can both take *sie* (she) and *es* (it). The double- and multiple-gender nouns are nouns with unstable gender. A double-gender noun may take agreements with more than one gender, with no different in meaning. An example would be "doctor", which can take "he" when it denotes a male or "she" when it denotes a female. Nouns like that are called nouns of common gender. While common nouns take two different types of agreement forms,

epicenes take only one, but they can refer to beings of either sex. They have grammatical gender and generally denote non-humans, but a few denote humans. In Portuguese, *pessoa* (person) takes feminine agreement, but can refer to a male or a female referent. In other words, epicenes are nouns that denote sexed beings, without differentiating them according to sex.

In the next section, an analysis on Portuguese will be addressed with the purpose of discussing some questions that are paramount for understanding gender in that language once it is the object of study of this research.

3.7 Gender in Portuguese

Mattoso Câmara Jr. (1970) argued all nouns are gender assigned in Portuguese, not only those that denote beings, but also those that denote things. According to him, masculine is non-marked and looks like a more general form. On the contrary, feminine is marked and indicates a certain specialization. The author claimed the opposition between masculine and feminine is not limited to *–o* and *–a* endings as in *lobo* (male wolf) and *loba* (female wolf) respectively. One should consider masculine can also be marked by *–e* ending as in *mestre* (male master) and by the null ending ∅ as in *autor* (male author), in opposition to *–a* as in *mestra* (female master) and in *autora* (female author) respectively. Once it would not be a very economic rule to have three forms for masculine (*–o, –e, and* ∅ endings), Mattoso Câmara Jr. (1970) defends the idea masculine is null marked (∅), while feminine is marked by *–a* ending. He classifies the *–o* and *–e* endings in masculine nouns as thematic vowels. An evidence for that lies in the fact masculine nouns does not necessarily refer to masculine gender, but it can also have a generic meaning. For example, in "*O lobo é*

um animal feroz." (The wolf is a ferocious animal.), *lobo* refers to both male and female wolfs, despite the *–o* ending. This would happen for all masculine nouns.

It is important to mention Mattoso Câmara Jr. (1970) excluded from his analysis oppositions like those between *homem* (man) / *mulher* (woman), *galo* (hen) / galinha (chicken) as, according to him, in these cases sex is indicated in the lexicon by either a particular noun or by derivational form respectively. And this is the reason why those nouns do not follow the rule in which masculine nouns are unmarked with ∅ morpheme and feminine nouns are marked with *–a* morpheme. He defends the idea gender is a formal rather than a semantic category in Portuguese.

Mattoso Câmara Jr. (1970) claimed gender inflection is a coherent mandatory mechanism in Portuguese. However, Villalva (1994) argued his analysis is incomplete and incoherent. She argues he did not include an explanation why the opposition for feminine nouns like in *casa* (house) and masculine nouns like in *caso* (case) does not have any semantic or morphological relation. Moreover, the examples of gender opposition that he gave are explainable by different etymology origins and not by gender variation. Finally, he did not explain why so many nouns do not have gender inflection in Portuguese.

Unlike Mattoso Câmara Jr (1970), Villalva (1994) defended the idea gender is not an inflectional category, since nouns such as *sobrecomuns*[7] and epicenes do not have gender contrast:

(72)	*(o) ídolo (idol),*	**(a) ídolo/a*
	(o) indivíduo (individual),	**(a) indivíduo/a*

[7] In Portuguese traditional grammar, epicenes and *sobrecomuns* both refer to nouns with grammatical gender without gender variation. But the former are used for animal names, while the latter are used for [+] human nouns. However, in this dissertation, both groups will be called epicenes since it is the conventional name in the international literature.

(73)	*(o) crocodilo (crocodile),*	**(a) crocodilo/a*
	(o) elefante (elephant),	**(a) elefante/a*
(74)	*(a) criança (child),*	**(o) criança/o*
	(a) pessoa (person),	**(o) pessoa/o*
(75)	(a) águia (eagle),	*(o) águia/o
	(a) cobra (snake),	*(a) cobra/o

In the examples above, nouns in (72) and (73) can only be assigned masculine gender, while nouns in (74) and (75) can only be assigned feminine gender. This way, Villalva (1994) concluded gender variation is not mandatory in Portuguese, because it does not affect all nouns. Consequently, it cannot be an inflectional property of all nouns. If gender were inflectional, it would behave like other inflectional categories such as number. However, gender can take a variety of forms and processes in the language. In Portuguese, not all words allow gender contrast as it was discussed above for epicenes. And words that allow gender contrast do so at different ways. Nouns with semantic gender can use composition as in "aluno" (male student) and "aluna" (female student) or a whole-form as in "homem" (man) or "mulher" (woman). They can also reach derivational processes such as "europeu" (male European) and "européia" (female European) or adjunction processes as "águia-macho" (male eagle) or "águia-fêmea" (female eagle).

Villalva (2012) defends words are morphological units, in which stems are highly projected followed by morphological and morphosyntactic specifications. This way, gender suffixes do not have anything to do with gender or number properties, but they are morphological specifications of the stem. For her, gender variation can be a lexical, a derivational or a compositional process.

In this section, some essential characteristics of Portuguese morphology were discussed. The next section will be focused in one of those special characteristics – the unmarkedness nature of the masculine gender. Examples in languages such as Spanish and French will also be analyzed in order to understand more about the use of unmarked forms. Next, the discussion will turn to agreement aspects in Portuguese.

3.8 Unmarked and default forms

Corbett (1991) reminds there may be constructions in which the target has to agree with a controller that is not specified for gender, as an infinitive clause, or when a choice of gender would force greater specificity than is possible or desirable for the speaker. For example, speakers may desire to refer to a child but be unable to select a gender agreement based on sex. Many languages solve this problem by using the regular gender form, which is often called *neutral agreement form* or *default agreement form*. However, neuter may not be the unmarked gender since almost all nouns denoting humans are masculine or feminine. Thus the choice of neutral agreement may be understood as the selection of the gender that is most appropriate in semantic terms.

In Spanish, *ello* in *Antes me gustaba mucho ir a los partidos de fútbol, pero todo ello ya no me interessa*. (I used to like to go to soccer matches before, but it does not interest me anymore) and *lo* in *Lo curioso de esa situación* (The curious thing about that situation) are neither masculine nor feminine, they are neuter terms. On the contrary, *el* in *El curioso* (the curious) could be mean either "the curious man" or the "the curious one". The masculine determiner would have a generic reading in the second meaning for *lo*. In Portuguese, neutral forms can also be found in *tudo*

(everything), *isto* (this), *isso* (that), *aquilo* (more distant than that). Moreover, the generic meaning can also be found for masculine gender as mentioned in the previous section. For example, in *O curioso sempre descobre algo* (The curious always figures out something), *o curioso* may mean "A man always figures out something" or "The curious one always figures out something".

In French, the masculine singular form of the adjective is the stem only as in *petit* (small), while the feminine singular is the stem plus *–e* in *pétite*, for the masculine plural, the *–s* marker is added to the stem as in *petits*, while the feminine is marked by *–e* plus *–s* as in *pétites*. Thus the masculine form is unmarked for gender in the most literal case. A similar situation would happen in Portuguese for some nouns with ∅ ending, as shown in the last section for the masculine form *autor* (author) and the feminine form *autora*, with the exception that in masculine plural the vowel *–e* would be needed in *autores* for phonetic reasons, in opposition to the feminine plural *autoras*.

Moreover, when the gender used to refer to pairs or larger groups is analyzed, it is possible to discover interesting semantics of gender in a given language. For instance, in French, *les Américains* (the Americans) is used to denote males or both males and females. This is one semantic justification for the use of a particular gender resolution for conjuncts of different genders. Thus masculine plural can denote semantic gender neutralization in French. The same would happen in Portuguese. For examples, the masculine form in the plural in *os americanos* (the Americans) is used to refer to males or both males and females; on the contrary, the feminine gender in the plural as in *as americanas* can only refer to females. In other words, since masculine gender (in the singular or in the plural) is unmarked in languages such as

Spanish, French, and Portuguese, it works as default gender, that is, it neutralizes semantic gender conflicts in some contexts of use.

3.9 Gender agreement in Portuguese

One can noticed a very good way to figure out gender assignment is through agreement. Corbett (1991) claims agreement is important for two reasons: it is the way gender is realized in language use and it is the basis for defining gender and establishing the number of genders in a given language.

> The term agreement commonly refers to some systematic covariance between a semantic or formal property of one element and a formal property of another. For example, adjectives may take some formal indication of the number and gender of the noun they modify. (STEELE 1978:610 apud CORBETT, 1991:105)

The common ways of marking agreement are in inflectional affixes, which are located before (prefixes) or after the stems (suffixes). Indo-European languages mark gender with suffixes. In Portuguese, for instance, the stem "gat-" can take *–o* for masculine (male cat) or *–a* for feminine (female cat). In Portuguese, agreement is shown in most nouns, determiners, adjectives, demonstratives, possessives, participles, relative pronouns, personal pronouns, adverbs, and complementizers. It is often said languages like Portuguese are rich inflected because of its redundant marking.

(76) *A[fem, sg] bailarina[fem, sg] admira[sg] os[mas, pl] psicólogos[mas, pl] que ajudaram[pl] ela[fem, sg] gentilmente depois de*

uma[fem, sg] das[fem, pl] fases [pl] mais intensas[fem, pl] de sua[fem, sg] vida[sg][8].

(The ballet dancer admires the therapist who gently helped her after on of the most intense phases of her life.)

In the sentence above, the determiner "A [fem, sg] ", the verb "admira[sg] ", and the pronoun "ela[fem, sg]" agree with the noun "bailarina[fem, sg]", while the determiner "os[mas, pl]" and the verb "ajudaram[pl]" agree with the noun psicólogos[mas, pl]. In the prepositional phrase "das fases mais intensas", the modified preposition "das (de+as)[fem, pl]" and the adjective "intensas[fem, pl]" agree with the noun "fases [pl]". The marks described in (76) are overt, but there are some covert marks in (76) too, as in the relative pronoun "que [masc, pl]" and in the noun "fases[fem]", for example. Thus the redundant agreement can be found, for instance, in the repetition of the information [fem, sg] in the determiner "A [fem, sg] ", in the noun "bailarina[fem, sg]", and in the pronoun "ela[fem, sg]".

Corbett (1991) affirms gender agreement can interact with tense, person, and number. In Portuguese, for example, there is a strong interaction between gender and number. The suffix *–a*, for example, marks feminine gender and singular number at the same time.

The next section will be focused on reviewing some empirical evidences of gender processing in Italian, Spanish, and Portuguese. This background knowledge on gender processing will be eventually useful for the present study.

[8] Adapted from one of the sentences used in the present study.

3.10 Previous research on gender processing

Romance languages are ideal to study gender morphology for at least two reasons: they have phonological gender systems, and in the majority of cases, their gender markers are overt. Consequently, a great number of studies on gender morphology were conducted in languages such as Italian, Spanish, and Portuguese. In this section, some of these studies will be discussed in order to clarify at which point gender surface cues can influence gender assigning and lexical retrieval.

3.10.1 Gender processing in Italian

In Italian, there are only two genders, masculine and feminine. Gender is an inherent property of every noun and gender agreement is marked on almost all modifiers, all pronouns and all past participles. There is no unmarked or zero noun forms in Italian. All nouns end in a vowel, and gender and number are marked together on that final vowel. Most masculine nouns end in *o* in the singular and *i* in the plural. These are considered phonologically transparent items. For a minority of masculine and feminine nouns, the final vowel is *e*. And because gender cannot be figured out from surface form alone on these words, they are called phonologically opaque. Moreover, a few words have contradictory marking, that is, their gender is the opposite of that conveyed in their morphology, for example, *la mano* (hand).

Bates et al. (1995) was interested in studying the role of gender in lexical retrieval in Italian. They authors used two types of lexical retrieval task: word repetition and gender monitoring (classification of nouns according to their gender). The first method is more automatic so that participants would not necessarily retrieve the gender of the nouns. The second method would require participants to retrieve the

gender of the nouns. The materials used in the experiments were composed as the following: out of the 468 nouns, half was feminine and half was masculine. 80% of those words were phonologically transparent (-*o* or –*a* endings) and 20% were phonologically opaque (-*e* ending). The authors also controlled for other factors, as for example, the number of syllables, semantic gender, humanness x abstractness, and frequency.

Bates et al. (1995) found out native speakers of Italian take less time to classify a word as masculine or feminine if that word has phonologically transparent morphology, compared to words with phonologically opaque morphology. In addition, they made more errors on nouns with opaque endings. Since there were no effects of gender in word repetition task, the authors concluded surface gender marking might be restricted to a later stage in processing, in which conscious attention to gender in required. This way, the authors defended there are two stages for gender morphology processing: in the pre-lexical phase, nouns may be accessed in the mental lexicon, probably under the rules described by the Cohort Model; then, in the post-lexical phase, surface gender morphology would play a role so that gender information would be checked before gender is assigned in tasks such as gender decision task. It should be mention factors such as the sex of the referents to which the nouns refer to, and the semantics of the nouns such as humanness, abstractness, and concreteness did not show any effects on the results.

Caffarra et al. (2015) aimed at investigating whether and when the sentence parser is sensitive to gender regularity of nominal endings in Italian. The authors conducted an ERP study on sentence comprehension using determiner-nouns pairs in embedded sentences. They manipulated agreement factors and gender-to-ending consistency, that is, phonologically transparent nouns (feminine nouns with –a ending

and masculine nouns with –o endings) and irregular nouns (masculine nouns with *–a* endings and feminine nouns with *–o* endings).

Caffarra et al. (2015) found main effects for gender-to-ending consistency and agreement, without interactions. During the first two windows at 350-500ms and 550-750ms respectively, transparent nouns evoked a larger frontal negativity than irregular nouns. In the third window at 750-950ms, a posterior positive response was elicited with higher amplitude for transparent nouns than for irregular nouns. In relation to the agreement factor, the disagreement condition elicited a more negative wave than the agreement condition at 350-500ms (LAN effects). Then a larger positivity was evoked for the disagreement condition than for the agreement condition at 550-750ms (P600) and at 750-950ms. It should be mentioned the positive effect at 550-750 did not reach statistical significance.

Caffarra et al. (2015) concluded the parser is sensitive to the gender morphological cues since they found differences between transparent and irregular nouns in Italian. This way, gender-to-ending consistency can be detected not only on isolated nouns in explicit gender decision tasks as in Bates et al. (1995), but also in sentence reading for comprehension. The findings of Caffarra et al. (2015) contradict Bates et al. (1995) since the latter only found effects of surface morphology at a post-lexical phase, while the former found effects of surface morphology at early ERP stages.

Caffarra et al. (2015) explained the lexical route might have priority in determining not only the noun's gender but also in establishing agreement dependencies, since the gender-related endings did not interact in the processes of repair and reanalysis. Furthermore, an explanation for the sustained effects of gender-to-gender consistency throughout the time windows is that morphological cues

conveyed by the noun ending are maintained online along processing in order to be integrated to sentence information in the end (*wrap-up effect*).

3.10.2 Gender processing in Spanish

In Spanish, most feminine nouns end in *–a*, and most masculine nouns end in *–o*. However, there are a number of opaque nouns ending ("*-e*", "*-n*", "*-l*", "*-s*", "*-j*", "*-r*", "*-z*") not clearly associated to a specific gender class. Moreover, there is a small group of words that reverse the typical gender-to-ending correspondence. For example, *mano* (hand) is feminine although it ends in *–o*, while *problema* (problem) is masculine although it ends in *–a*.

Afonso et al. (2013) were interested in investigating to which extent the lexical information of gendered-marked nouns would be used in a gender decision task. In Experiment 1, the authors manipulated the nouns endings, which could be regular (–a / –o) or gender-correlate (–ón is more frequent for masculine and –ad is more frequent for feminine). The results indicated faster reaction times and higher accuracy were found for regular ending forms, which seems that the morphology influences in gender decision process. In Experiment 2, the authors used a masked priming task in which definite articles (*la* / *lo*) or possessive pronouns (*mi* / *tu*) would prime for marked (*–a* or *–o* endings) and opaque nouns. The results showed marked nouns (with *–a* or *–o* endings) had faster gender decision times than other nouns. Moreover, it seems the information in the definite article was only relevant for gender access of unmarked nouns. Therefore, gender assignment in Spanish seems to take a dual-route mechanism for gender access: gender-marked nouns are accessed based on

their morphology cues, while unmarked nouns would require the retrieval of the gender conveyed in the definite article.

Caffarra et al. (2014) also provided evidence for existing two routes for gender retrieval in Spanish. They conducted in ERP study with article-noun word pairs. The nouns used in the experiment could be either transparent (*–a* ending for feminine nouns and *–o* ending for masculine nouns) or opaque (other endings). The behavioral results showed the participants were faster and more accurate in the agreement than in the disagreement condition. Moreover, transparent nouns were responsible for higher accuracies than opaque nouns. The ERP results indicate the disagreement condition elicited a greater negativity than the agreement condition at the 350-500ms and 500-750ms windows. And in the case of the disagreement condition, transparent nouns showed greater negative amplitudes compared to opaque nouns at the second window.

This way, there was no interaction of transparency and agreement effects in the first window, which means the initial computation does not rely on surface morphology. Caffarra et al. (2014) explained the initial effects of gender mismatch between the determiners and the nouns might reflect some abstract features specification stored in the lexicon. The authors concluded for transparent nouns, two different routes were available (the lexical and the form-based), whereas for opaque nouns, only the lexical route can be used. Furthermore, it is worth noting the presence of formal cues to gender affects left hemisphere earlier than right hemisphere.

3.10.3 Gender processing in Portuguese

Corrêa et al. (2004) investigated whether gender inflected nouns and adjectives are represented as full forms in the mental lexicon in Portuguese. The

authors conducted a series of lexical decision tasks manipulating the gender (masculine or feminine), the frequency dominancy[9], and the grammatical category (noun or adjective) and controlling for the cumulative frequency[10].

In Experiment 1, Corrêa et al. (2004) tested whether the frequency dominance would affect the speed of recognition of nouns and adjectives. The authors found effects of frequency-dominance for feminine nouns and for masculine adjectives, but they only found a mean tendency for the latter. This finding led to the conclusion feminine dominant gender is fully-lexically accessed, since the these nouns may be recognized by their surface and not my their morphemes. An explanation for that is the following: feminine gender may imply a modified meaning in the noun stem, which is configured as a derivational process. Experiment 2 was similar to Experiment 1, except for the fact that it only contained nouns. The results replicated Experiment 1, that is, feminine dominance impacted lexical access of nouns. In Experiment 3, only masculine nouns were used. The results indicated masculine nouns were not influenced by frequency-dominance. A possible reason for that lies in the fact that masculine forms are no gender inflected, they are unmarked for gender. Experiment 4 contained only feminine nouns and adjectives. The results replicated Experiment 1 since frequency dominance affected the recognition of feminine nouns and adjectives, especially the former.

Corrêa et al. (2004) claimed only feminine dominant nouns are fully represented in the mental lexicon. Moreover, adjectives are recognized in the function of their masculine unmarked form. Thus it seems nouns and adjectives are represented and accessed at different ways. Gendered-inflected nouns may require additional

[9] Frequency dominancy is the relative frequency of the surface form of an inflected pair. For example, if the singular use is more frequent than the plural use for a particular pair of words, it means that this word is singular-dominant.

[10] Cumulative frequency is the sum of the frequency of each member in a pair of words.

semantic features for the feminine form, and since feminine dominant nouns are few, they would be an exception as they might have an independent representation for them in the mental lexicon. On the contrary, adjectives do not imply a change of meaning in the stem; therefore, they would be syntactically accessed on the basis of the masculine unmarked form, similarly to masculine nouns.

Resende (2015) was interested in investigating the role of morphology in gender retrieval in Brazilian Portuguese. Similarly to Spanish, in Portuguese, most nouns ending in *–a* are feminine, while most nouns ending in *–o* are masculine. However, there are nouns with other endings that can be either feminine or masculine, just like Spanish and Italian. On the one hand, nouns ending in *–agem* and *–ade* are always feminine as in *garagem* (garage) and *imagem* (image). On the other hand, words ending in *–or* and *–ema* are always masculine as, for example, *amor* (love) and *problema* (problem). In addition, there are words with –e ending that do not provide any cue for gender, as for instance, the masculine noun *leite* (milk) and the feminine noun *noite* (night). Finally, like in Spanish and Italian, there are nouns that contradict the phonological form, that is, masculine nouns that end in *–a* and feminine nouns that end in *–o* as in *[o] dia* (day) and *[a] mão* (hand).

Resende (2015) conducted 3 behavioral experiments in Brazilian Portuguese. In Experiment 1, determinant-noun pairs were contrasted into regular (*-agem* and *-ade* endings for feminine nouns, and *–ume*, *-une*, *-or* for masculine nouns) and opaque (-e endings) in agreement and disagreement conditions. She also controlled for the frequency of the nouns. The results indicated frequency effects for both regular and opaque forms, but larger for regular ones. The author concluded there was evidence in favor of a single-route mechanism, that is, full access in memory, since both regular and opaque forms were affected by frequency.

Experiment 2 reported in Resende (2015) was a gender-assigning task with inanimate nouns in regular, opaque, transparent (feminine nouns ending in *–a* and masculine nouns ending in *–o*), and irregular (masculine nouns ending in –a, and feminine nouns ending in –o) forms. It was found faster reaction times for regular>transparent>opaque>irregular forms in this order. She explained Experiment 2 replicated Experiment 1, since frequency was the key effect to explain the results of the reaction times. Experiment 3 tested pseudo-words and verified the more the pseudo-words assemble real words, the easier is lexical retrieval, and the faster is the reaction time in the gender-assigning task. Again, the author found evidences in favor of a single-route mechanism to explain gender lexical access.

Resende (2015) also found out a connexionist simulation of a system with a single-route mechanism system could replicate the behavior of the participants in Experiment 3. Thus she argued gender decisions were only based on associative relations among structures that share the same features in the mental lexicon.

Finally, in her last experiment, an ERP study, Resende (2015) contrasted agreement and disagreement determinant-noun pairs (condition 1) and noun-adjective pairs (condition 2) with regular and irregular gender forms. The results suggest condition 1 evoked a biphasic LAN/P600 for both forms, while condition 2 only elicited a P600 for regular forms. Therefore, Resende (2015) concluded since both regular and irregular forms evoked the same ERP effects, it seems they share the same neurocognitive mechanism for agreement between determiners and nouns. However, the agreement relation between nouns and adjectives is different in the sense that only a reanalysis effect of P600 was found.

3.10.4 Summary

Taken together, the works of Bates et al. (1995), Caffara et al. (2014, 2015), Afonso et al. (2013), Corrêa et al. (2004), and Resende (2015) showed nouns can be retrieved at different ways depending on their gender surface cues. They are all in favor of a dual-route mechanism for gender access for Italian, Spanish, and Portuguese, except for Resende (2015), who argues in favor of a single-route mechanism for Portuguese.

With respect to Italian, it is important to mention Caffarra et al. (2015) found different results from Bates et al. (1995). Bates et al. (1995) showed morphological cues are only detected at a post-lexical stage with the purpose of feature checking, while Caffarra et al. (2015) showed morphological cues are early detected by the parser, and continue to influence processing until the end of the sentence.

For Spanish, the findings of Caffarra et al. (2014) are in line with Pinker (1991) and Allen et al. (2003). The authors argued regular forms (phonologically transparent) needed to be decomposed by mental grammar, while irregular forms (phonologically opaque) were fully accessed in the mental lexicon. It is worth mentioning Afonso et al. (2013) also proposed that irregular forms might also depend on the context (definite articles) to retrieve gender information.

Finally, evidences for Portuguese are contradicting. Corrêa et al. (2004) found strong evidences that feminine dominant nouns are fully accessed in the mental lexicon, while masculine nouns and masculine and feminine adjectives may be morphologically parsed. Resende (2015) claimed it was the mechanism of memory association in the mental lexicon that was able to explain gender retrieval. Thus more studies on gender processing are needed in other to reach a conclusion for Portuguese.

3.11 Different types of gender

According to Corbett (1991), gender distinctions are crucial for every living being, and that is the reason why some languages developed semantic gender, in which gender is congruent with sex. For example, most languages have different words for "woman" and "man", *mulher* and *homem* in Portuguese; or different pronouns for female and male referents as "he" and "she", *ele* and *ela* in Portuguese. Gender variation in semantic gender can be expressed by either whole-forms or derivational/compositional processes. In the former case, words must have gender information defined in their lexical representation. In other words, gender information is part of the definition of the word. This type of gender is called definitional gender (for example, "man" and "woman", *homem* and *mulher* respectively in Portuguese). In the latter case, gender variation requires compositional/derivational processes as in *menino* (boy) and *menin-a* (girl) in Portuguese or "actor" and "actr-ess" in English.

Furthermore, languages also differ in the way they include gender distinctions to entities without sex. For example, in Romance languages, all nouns are marked for gender – masculine or feminine. However, in the case of inanimate nouns, animal names, and some nouns that refer to humans, gender is not semantically motivated; it seems to be only a property of lexical items. It is said languages like that have grammatical gender.

Pragmatic factors related to world knowledge can also influence gender marking. Role nouns and occupations are generally gender stereotyped towards feminine or masculine. In this case, gender information can be extracted from a probabilistic view based on world knowledge. For example, words like "nurse" are stereotypically feminine because in the real world there are more female than male nurses.

As one can see, gender can be influenced by semantic, lexical, grammatical, and pragmatic properties. Studies on gender can be a fruitful way to understand how languages actually work, in their representation and processing. This section will be focused on reviewing and discussing some important studies on different types of gender. Although those works are on language processing, theoretical and linguistic representational issues will also be addressed.

3.11.1 Is grammatical gender always arbitrary?

Vigliocco et al. (2005) tested two hypotheses intended to explain how a child would learn gender: the first hypothesis is gender effects arises as a consequence of similarity in linguistic contexts; the second hypothesis is gender arises as a generalization from the transparent relationship between sex of human referents and gender of nouns.

The first hypothesis comes from the basic idea words that have similar syntactic and morpho-phonological properties also tend to have similar meanings. This is known as syntax-to-meaning mappings. Thus, nouns that share the same gender are used in the same linguistic contexts, which differ from other contexts in which nouns of a different gender are used. The *similarity and gender hypothesis* can be tested in languages like Italian and German because as they are morphologically rich languages, they can provide a good number of gender-marked sentential contexts.

The second hypothesis is based on associations between gender and sex, which is a consequence of co-occurrence of linguistic features and conceptual features. According to this hypothesis, learning gender in Romance languages would be easier than leaning gender in German, for example, because it would be harder for

a child to establish the association between sex and gender in a language with three genders like German.

Vigliocco et al. (2005) advocated speakers of different languages may pay more attention to what is obligatorily expressed in their language rather than to what is optional. One of the mandatory mechanisms in Italian and German is grammatical gender. The relation between grammatical gender and conceptual properties of objects seems to be very arbitrary. Taken this into account, the authors carried a series of experiments and Italian, German, and English (baseline as it does not have grammatical gender) in order to test for the effects of meaning similarity among words with grammatical gender referring to animals and objects. In Experiment 1 speakers of Italian were presented to three words, and they were instructed to judge which two of the three were most similar in meanings. If the first hypothesis were true, gender effects would be found for both animals and objects; however, if the second hypothesis were true, gender effects would only be found for animals. In Experiment 2 speakers were instructed to name pictures aloud under time pressure in order to test the hypotheses in a more automatic *on-line* manner.

The results of Vigliocco et al. (2005) are the following: gender-to-meaning effects were limited to Italian (two-gender system); and in Experiment 2, only for animals, and not for artifacts. It should be noticed there were no phonological overlap effects. Thus the authors concluded the second hypothesis (*gender to sex hypothesis*) was better appropriate to explain their results. In languages such as Italian, speakers might pay more attention to the sex of referents in order to produce correct words, which is in contrast with English as conceptual gender is less obligatorily marked in this language. It seems grammatical gender is not arbitrary, as theory has affirmed

when it refers to sexed beings, like animals. On the contrary, in these cases, grammatical gender is influenced by semantics related to the referents meaning.

Konishi (1993) also studied the semantics of grammatical gender. He asked Spanish and German speakers to rate words on a semantic scale, and found out nouns with masculine grammatical gender that have masculine connotations as words like "power" were rated higher on semantic dimensions. The same was not found for feminine gender. Moreover, words with different grammatical gender across the languages were also rated differently, as "fork", which is masculine in Spanish (*tenedor*), but feminine in German (*gabel*). Since Konishi (1993) only tested for inanimate nouns, *the sex and gender hypothesis* of Vigliocco et al. (2005) could not be tested in his study.

Corrêa (2001) investigated the role of semantic context in how epicenes are retrieved by pronouns in Brazilian Portuguese. According to her, despite having an intrinsic gender, epicenes were more easily retrieved when accompanied by a semantic gender-matching context. However, it should be highlighted semantic context was also taken into account at late processing phases.

To sum up, it seems grammatical gender can also be influenced by semantics, especially when referring to sexed beings. This is evidence in favor of the idea that gender is always semantic motivated (Corbett, 1991).

3.11.2 Comparing grammatical gender and semantic gender

Casado et al. (2017) was interested in examining whether the sex of the agents involved in the communication act (speaker and listener) would influence the activation and selection of gendered words in Spanish words, which could be nouns

with semantic gender (when the sex of the referent coincides with the gender of the word) or with grammatical gender (arbitrarily gendered words without information regarding the sex of the referent). They used 3 kinds of tasks in their study, word repetition, lexical decision, and gender decision. The authors only used transparent gender marked nouns in their experiments, that is, masculine nouns ending in –o and feminine nouns ending in –a. The results point out that in the 3 tasks, semantically gendered nouns were retrieved faster than arbitrarily gendered ones, which could reflect an effect of animacy, since the semantically gendered nouns were all animate, while the arbitrarily gendered nouns used by Casado et al. (2017) were inanimate.

The effects of the sex of the participant were only found in the third task of Casado et al. (2017). When the sex of the participant matched the gender of the nouns, retrieval was facilitated. It should be noticed that this effect occurred in both semantically and arbitrarily gendered nouns. The authors explain that people encode and organize information according to their-selves sex role. This way, words related to their self-sex are more frequently used, which generates a priming effect.

Furthermore, the effects of the sex of the speaker were found in the lexical decision task and in the gender decision task. When the sex of the speakers matched the gender of the words, facilitation was perceived (in the lexical decision task, only for semantically gendered masculine nouns; and in the gender decision task, only for semantically gendered feminine nouns). Finally, taken together, these results, according to Casado et al. (2017), suggested that both female and male participants included female and male representations when hearing semantically gendered masculine nouns, which is evidence in favor of the fact that masculine works as the generic or default gender in Spanish. In other words, when a male speaker uses a masculine gendered word, both male and female listeners would think on either the

male or the female referents, but the strength of activation would be drawn to the male referent. And that would be easier for a male listener than for a female listener. However, when a male speaker uses a feminine gendered word, it would be hard for listeners to think on either male or female referents, especially if the listener is male. The final conclusion of the authors is that despite gender information be stored in the lexico-syntactic level, nouns could were processed in a top-down manner in their study, that is, from semantic-pragmatic information to lexico-syntactic level.

Vigliocco & Franck (1999) were interested in determining whether Italian and French agreement production would be sensitive to the distinction between grammatical and semantic gender. The authors hypothesized for semantic gender there would be a match between the syntactic and conceptual gender, while for grammatical gender, there would be gender specified by syntactic features only. Thus, they predicted more errors of agreement for nouns with grammatical gender than for nouns with semantic gender. In other words, semantic information would help correct agreement since it provides redundant compatible information.

Vigliocco & Franck (1999) conducted a series of four experiments eliciting gender agreement errors between subjects and predicative adjectives. The participants were presented to an adjective and then a sentential fragment. They were instructed to repeat the fragment and complete it with the adjective informed. All fragments were composed by subject head noun followed by an embedded modifier prepositional phrase (local noun).

As expected, Vigliocco & Franck (1999) reported more gender agreement errors between subjects and predicative adjectives for subject head nouns with grammatical gender than with semantic gender in Italian and French. Moreover, gender mismatch between the subject head noun and the prepositional head noun also

caused more agreement errors. Feminine subject head nouns were also responsible for more agreement errors than masculine in French, but this difference was not found in Italian. It is worth mentioning there were no animacy effects in the results, there is animate nouns (referring to humans or animals) were not treated differently from inanimate nouns (objects).

Vigliocco & Franck (1999) believed agreement errors were caused by the interference of the gender features of the local noun on the gender features of the subject head noun. This way, the gender features of the local nouns were erroneously taken as the features of the subject. The asymmetry between feminine and masculine genders in French might be explained by the fact that feminine gender in French is marked by the insertion of *–e* morpheme to the noun stem, which might be more costly. This way, more errors were found in this condition. The idea of justifying the gender asymmetry with feminine markedness was discarded because the majority of errors were found for grammatical gender, which may not have the marked/markedness dichotomy. Finally, less errors for nouns with semantic gender reveals redundant information ensures accuracy and allows a more efficient encoding.

Cacciari et al. (1997) urged languages that have grammatical gender such as Romance languages could be advantageous to examine the role of gender cues in pronoun resolution. The reason is they can disentangle morphosyntactic gender from semantic gender. The aim of Cacciari et al. (1997) was to verify whether, how, and when surface information, more precisely, gender cues, influences pronoun resolution. They conducted self-paced reading experiments to test two types of nouns in Italian, (a) epicenes, which are nouns with grammatical gender, such as *la vittima* (the victim), in which morphosyntactic gender information is marked both in the word ending and in the article; and (b) ungendered words, which can be assigned both

masculine and feminine semantic genders, such as *l'amante* (the lover). In this case, gender is provided neither by the word ending nor by the article as in Italian there is a contraction between definite articles and nouns initiated by vowels.

The results in Cacciari et al. (1997) indicated pronoun resolution was facilitated by epicenes in comparison to ungendered nouns, which means surface gender cues are taken into account in pronoun resolution. When the pronoun matches the surface cues of the antecedents, coreference is facilitated. They concluded morphosyntactic gender is relevant to pronoun resolution. The authors emphasized they did not find effects of more costly processing in pronouns that mismatched the grammatical gender of the epicenes, which means readers may have either a masculine and a feminine referent associated to an epicene noun in their mental model.

Lawall et al. (2012) compared epicenes with bigender nouns, which have a variable semantic gender in the language. The authors conducted a self-paced reading on pronouns processing manipulating the type of antecedent, which could be an epicene or a bigender. They predicted to find shorter reading times at the pronoun for sentences in which the antecedents agreed in gender with the pronouns. The results found by Lawall et al. (2012) indicated the participants were only sensitive to lack of agreement between epicenes and pronouns in an off-line phase. The participants did not accept well a conceptual coreference with a gender mismatching epicene. Furthermore, masculine pronouns referred more easily to a feminine epicene than feminine pronouns to a masculine epicene in the on-line results. The authors explained that in this case, since masculine gender is default, it could refer to either a masculine or a feminine antecedent.

Alves (2014) also compared epicene and bigender antecedents in pronominal resolution. The hypothesis was the following: epicene antecedents would be more easily processed than bigender antecedents because, unlike the latter, the former have fixed gender; consequently, they would not need to rely on context. Moreover, the expectation was masculine pronouns would also refer to either masculine or feminine antecedents. It was also predicted that gender matching between antecedents and pronouns would facilitate coreference processing. The results confirmed the predictions. In addition, it was found masculine epicene and bigender antecedents being retrieved by feminine pronouns were more accepted by female than by male participants in the off-line results. However, feminine epicenes in masculine biased contexts were more accepted by male participants than by female participants. This way, it seems male participants are masculine biased. These results are in congruence with the ones found in Italian by Cacciari et al. (2011).

3.11.3 Understanding stereotypical gender

According to Oakhill et al. (2005), background knowledge influences the construction of the information in a text providing cognitive economy. The authors studied one specific type of background knowledge: stereotypical gender information. Psycholinguists often use the following riddle to exemplify stereotypical gender:

> This morning a father and his son were driving along the highway to work, when they were involved in a horrible accident. The father was killed, and the son was quickly driven to the hospital, severely injured. When the boy was taken into the operating theatre, the surgeon exclaimed: “Oh, my God, this is my son!

The riddle is based on the fact that when there is not any specific information about the gender associated with an occupation or role noun, as “surgeon”, prior knowledge, in this case, gender stereotype, would be used by the reader to infer the more likely gender, which in this case is masculine. However, once that inference is proved incorrect by the end of the story, readers would have to reconstruct their character representation, which is more costly for processing.

In English, most terms for occupations are not gender marked, and many of them are gender stereotyped. So, if a reader supposes a surgeon is male and a secretary is female, it is so due to real statistics and world knowledge. However, a big question in the literature is whether the stereotyped information is immediately activated by the time a stereotyped word is encountered, or only later, when required by discourse, as for example, when a pronoun needs to be resolved.

Oakhill et al. (2005) conducted a series of experiments in English to check whether gender stereotypical information is automatic and unable to be suppressed. In their experiments, the participants were instructed to decide whether or not two terms (an occupation and a kinship term) could refer to the same person (e.g.: nurse-aunt; nurse-uncle).

The authors found out people were biased by gender stereotypes. For example, participants’ made more errors or had longer reaction times to accept pairs like “engineer-mother” than “engineer-father” or with gender-neutral terms. This way, both accuracy and reaction time suffered when stereotypical gender conflicted with the gender of the kinship term. Since the authors did not find facilitation effects of the congruent pairs, they affirmed stereotyped information is a case of interference. Moreover, Oakhill et al. (2005) reinforced gender stereotype might have a very strong and persistent effect because participants could not suppress it even after being

strongly warned about it. The authors concluded gender stereotype associated with occupations and roles is incorporated into the information representation immediately as soon as an occupation or a role name is read. In other words, it is automatic.

Carreiras et (1996) argued when reading, a representation of the situation is incrementally formed in our mental model through text input, context, and world knowledge. A particular type of world knowledge is gender stereotype. However, the authors alerted that, unlike English, in Spanish, stereotype information is not used to infer a person's gender because gender is explicitly given by the definite article. In order to compare how stereotypical gender influences comprehension in those two languages, self-paced reading was used to present short texts with a stereotypical role name followed by a pronoun (he/she).

(77) Examples of the materials in Carreiras et al. (1996)
a. The footballer wanted to play in the match.
He/She had been training very hard during the week.
b. El futebolista quería jugar el partido.
El había estado entrenando mucho durante la semana.
c. La futebolista quería jugar el partido.
Ella había estado entrenando mucho durante la semana.
d. El cantante/La cantante recibió una gran ovación.
(The singer received a great ovacion)
El/Ella tuvo que salir de nuevo al escenario.
(He/She had to appear on stage again.)

On the one hand, the results in Carreiras et (1996) showed reading times for the second sentences in English were longer when there was a mismatch between the gender of the pronoun and the stereotyped gender of the role noun. On the other hand, the results for Spanish showed reading times for the first sentences were longer when

there was a mismatch between the gender of definite article and the gender of the stereotyped role noun.

Carreiras et (1996) concluded stereotypes can influence the representation for a particular character in a text as soon as they become available. When they are the only source of gender information (as in the first sentences of the examples in English) they can be used to infer the gender of the characters in the mental model. However, if some upcoming information that mismatches the stereotyped inference appears in the text, that previous inference must be overridden, which might explain the difficulties the readers had in these cases. Although gender assignment in Spanish does not require a stereotyped inference as in English once gender information is already given in the articles, readers may also have to encode something new in their mental models: the clash between the gender of the articles and the stereotypical gender of the characters.

Kennison & Trofe (2003) also examined the role of gender-stereotyped nouns during language comprehension. First, they conducted a rating study with 405 nouns and compound nouns. Out of these nouns, 32 words were strongly stereotyped to refer to males and 32 to females. Thus these 64 nouns were used in a self-paced reading study on pronoun resolution, but unlike Carreiras et al. (1996), in which the participants saw the whole sentence, Kennison & Trofe (2003) used the phrase-by-phrase moving window paradigm. This way, they believed they would obtain more information about the time course of information than Carreiras et al. (1996).

In the rating study, Kennison & Trofe (2003) found out male participants gave the nouns more masculine ratings than female participants. This might mean male participants view the world as composed by more males than females, while female participants view the world as composed by more females than males. With respect to

the self-paced reading study, reading times were longer on two regions after the pronouns when the gender stereotype of the antecedent mismatched the gender of the pronoun than when they matched. This result replicated Carreiras et al. (1996). Moreover, reading times at the pronoun "she" were longer than at the pronoun "he". The authors believed it was because "he" is more frequent than "she". It should be noted the factor "sex of the participants" had no effects on the results of the self-paced reading experiment.

Kennison & Trofe (2003) concluded word-specific gender stereotypes influences resolution of pronouns. They hypothesized comprehenders must have a representation for each word stored in memory, and this representation includes information about the word's gender stereotype, which is computed according to the relative frequency of usage in referring to males and females.

Cacciari & Padovani (2007) conducted two priming experiments aiming at investigating gender stereotypes in Italian. In Experiment 1, participants were shown pairs of stimuli formed by a prime word and a target gender-marked pronoun. The participants were asked to decide the grammatical gender of the pronoun ignoring the prime word. The primes were role nouns with and without gender stereotypes, and the stereotyped gender of the role nouns would match or mismatch the gender of the pronouns. No statistically significant priming results were found in Experiment1. Thus, the authors decided to give the participants more time to process the prime in Experiment 2 so that the stimulus presentation time increased from 200 to 300ms and the interval between the prime and the target increased from 100 to 200ms.

The results of Experiment 2 in Cacciari & Padovani (2007) showed reaction times were faster when the gender of the pronoun, matched the stereotyped gender of the role noun. Thus although it is less stable than semantic gender, the authors

concluded stereotypical gender might be part of the nouns's mental representation. In addition, the results indicated when feminine stereotyped role nouns mismatched the pronouns (teacher-he), an inhibition effect was found in the response. However, no inhibition effect was found for masculine stereotyped nouns that mismatched the gender of the pronouns (engineer-she). The authors stated since masculine is unmarked in Italian, it might spread activation to conceptually masculine and feminine units, while the marked feminine would activate only the feminine unit.

Duff & Keir (2004) analyzed whether processing of individual words could be modified by discourse context. Their first goal was to show how gender stereotypes are accessed during comprehension and how they interfere in processing when stereotypes are violated. Their second goal was to investigate whether such interference effect could be reduced by discourse context.

In Experiment 1, Duff & Keir (2004) monitored the participants' eye movements while reading sentences containing role nouns or occupation terms followed by reflexives, which could match or mismatch the stereotypical gender of the antecedents. In Experiment 2, the authors created a series of paragraphs focused on a particular character identified with a stereotyped role or occupation. In the disambiguating condition, the gender of the character was explicitly stated before the character was introduced informing the participants whether the character was a woman or a man. On the other hand, in the neutral condition, the gender of the character was not stated.

Duff & Keir (2004) found out in Experiment 1, and in the neutral condition of Experiment 2 reading processing suffered interference when text and the gender-stereotyped information conflicted (mismatch-cost). This could possibly mean gender stereotypes are immediately activated by the time the role names were found,

resulting in a conflict when the reflexives were subsequently encountered, or that gender information was not activated until the reflexives were encountered. Their study could not disentangle between these two possibilities.

It should be highlighted there was no interference in the disambiguating condition in Experiment 2, that is, when prior context specified the gender of the role names. Duff & Keir (2004) concluded gender stereotypes are automatic, but they can be overridden, at least for a short time, by early explicit specification of gender in the discursive context.

3.11.4 Comparing stereotypical gender and definitional gender

Kreiner et al. (2008) were concerned with the nature of the stereotypical gender. According to the mental models approach, stereotypical gender is inferred from world knowledge, while according to the lexical view it is stored in the lexicon as part of the lexical representation. A good way to disentangle between these two proposals is comparing stereotypical gender with definitional gender. Both lexical and inferential approaches predict processing difficulty in mismatching conditions, that is, when stereotypical or definitional genders mismatch the gender of the reflexives. However, according to the lexical account, there may be some quantitative probabilistic differences between definitional and stereotypical nouns. Definitional gender is more polarized, or even categorical, whereas stereotypical gender is more graded biased. This way, a larger mismatch-cost would be expected for definitional gender than for stereotypical gender. On the other hand, according to the inferential view, definitional gender and stereotypical gender are qualitative different, that is, the former might be

specified in the lexicon, while the latter might be a result of probabilities based on world knowledge.

Kneiner et al. (2008) showed the results of two eye-tracking experiments differed in function of the order by which the gender information is given in discourse. In Experiment 1, the role nouns were presented earlier than the reflexives (anaphora), while in Experiment 2, the reflexives were presented earlier than the role nouns (cataphora). In Experiment 1, the mismatch-cost were similar for both stereotypical and definitional gender. However, in Experiment 2, the mismatch-cost was only found for definitional gender nouns. The authors concluded, unlike definitional gender, stereotypical gender could be overridden by prior discourse. This way, there is a qualitative difference between definitional gender and stereotypical gender, which is evidence in favor of the inferential approach. The gender of definitional nouns might be lexically represented, while stereotypical gender might be pragmatically inferred. Finally, Kreiner et al. (2008) argued in favor of an interactive model in which lexical, pragmatic, and syntactic representations would efficiently communicate throughout processing allowing readers to incrementally integrate information from different linguistic levels.

Osterhout et al. (1997) were concerned about the mental representation and processes underlying stereotype-violating reflexives compared to reflexives that are consistent with the stereotypes. In addition, they were concerned whether the stereotype violation in an ERP study would resemble to pragmatically implausible words (N400) or syntactically anomalous words (P600). In the second scenario, a P600 brain response would be similar to the brain response elicited by definitional male or female nouns whose grammatical features are associated with the word's lexical representation.

The materials used for reflexive antecedents in Osterhout et al. (1997) were nouns specifying occupations (as words like "actress"), states (as words like "bachelor") and titles (as words like "duke"). The first type of nouns was gender-stereotyped, while the latter two types were definitional male or female nouns. Half of the definitional nouns were male (as "king") or female (as "queen"), and half of the stereotypical nouns were female stereotyped (as "babysitter") or male stereotyped (as "pilot").

Osterhout et al. (1997) reported violations of definitional gender evoked a larger P600 than violations of gender stereotypes. Moreover, the positive wave of P600 evoked for violations of gender stereotypes persisted even when the participants judged the sentences to be acceptable. The fact that both types of violations evoked P600 effects indicate the lexical representation of definitional and stereotypical nouns is similar, and they participate in the grammatical rules requiring agreement, that is, their gender information is encoded within grammar.

According to Osterhout et al. (1997), the different amplitudes of P600 between violations of definitional and stereotypical gender might suggest participants experienced more difficulties from recovering from a definitional gender violation than a stereotypical gender violation. In other words, violations with gender definitions result in an unavoidable ungrammaticality, whereas violations with gender stereotypes result in reanalysis. It seems initially, readers assign the stereotypical reading, but by the time they encounter the reflexives and they realize their first reading was not adequate, they are ultimately forced to assign the less preferred gender feature.

Osterhout at al (1997) also reported female participants showed a larger P600 than male participants in mismatch conditions for either a definitional or a

stereotypical antecedent. This can either mean females are better than males at detecting violations or females respond more strongly than males to violations of social expectations about "appropriate" gender roles.

Canal et al. (2015) aimed at replicating the results found by Osterhout et al. (1997). But more importantly, inspired by the different results found by Osterhout et al. (1997) regarding the sex the participants, they were interested in testing the hypothesis that the flexibility of the gender representation of a role noun might depend on the individual's social perception of gender.

Unlike Osterhout et al. (1997), Canal et al. (2015) claimed the difference between definitional and stereotypical nouns is not quantitative. Instead of finding P600 effects in the mismatch conditions for both types of antecedents as Osterhout et al. (1997), Canal et al. (2015) found out when the gender of the reflexives mismatched the stereotypical gender of the antecedents, the ERP results were biphasic showing negative effects in frontal left electrodes (Nref) and a positive effect in parietal electrodes (P600). It is worth mentioning like Osterhout et al. (1997), Canal et al. (2015) also found P600 effects when the gender of the reflexives mismatched the gender of the definitional antecedents. Based on other works, Canal et al. (2015) interpreted the Nref effect reflecting a search for additional information to link the reflexives to the stereotypical antecedents. In other words, as stereotypical gender does not determine gender categorically, but it is based on a probabilistic bias, comprehenders might need to look for additional information to realize antecedents and reflexives are coreferential.

Canal et al. (2015) did not find strong results indicating male and female participants process gender representation differently. However, they found out in mismatching conditions, participants who described themselves as less feminine

showed a larger negative response for stereotypical antecedents, and a larger P600 for definitional antecedents. On the other hand, male participants with more feminine traits showed a reduced P600 effect, but an increased P600 effect if participants were female. In a nutshell, less feminine or less sexist participants may have actively searched for an appropriate although less likely antecedent.

Canal et al. (2015) concluded language comprehension is influenced by larger non-linguistic factors such as the individuals' experience and personal beliefs. They explained the difference between their results and Osterhout et al. (1997)' may be linked to a couple of factors: (a) Osterhout et al. (1997)' used less electrodes, which might indicate they could have missed some effects; (b) the materials in Osterhout et al. (1997) were more stereotypically biased due to the use of modifiers; or (c) a social change - today's society is more liberal than 15 years ago.

3.11.5 Summary

Gender can be influenced by semantic, lexical, grammatical, and pragmatic factors. Due to this diverse nature, there are different types of gender across the languages. In Portuguese, one can find, among others, semantic gender, grammatical gender, definitional gender, and stereotypical gender.

In semantic gender, the gender of the words is congruent to the sex of the referents. According to Vigliocco et al. (1999), semantic gender facilitates gender retrieval due to redundancy as this type of gender is both conceptually and syntactically specified. Semantic gender can be expressed by a whole-form or compositional/derivational processes. In the first case, gender information is specified in the lexicon at the word's entry as in words like *mulher* (woman) and *homem* (man).

Nouns like those have definitional gender. In the second case, gender information is retrieved by the suffixes or compositional morphemes of the words as for example in words like *menino* (boy) and *menin-a* (girl). In this case, mental grammar/parser would need to decompose the words in order to have access to its morphemes, and this is more demanding than retrieving gender from definitional nouns, which are probably fully accessed in the mental lexicon.

Grammatical gender is the gender of inanimate nouns, abstract entities, animal names, and a few nouns referring to humans. Nouns that carry grammatical gender are called epicenes. According to Cacciari et al. (1997), epicenes may carry two mental representations of the referents, a female and a male one. Grammatical gender is invariable and considered as arbitrary in the language. However, Vigliocco et al. (2005) showed epicenes that refer to animals are semantically motivated because animals are sexed beings; consequently, *the gender to sex rule* typical of semantic gender can be applied. Moreover, it seems the grammatical gender of some abstract entities can also be semantically motivated as shown by Konishi (1993). One question left is whether epicenes that refer to humans are also semantically motivated. If so, to which extent they can be compared to nouns with semantic gender.

According to Vigliocco et al. (1999), grammatical gender and semantic gender are different types of gender. Grammatical gender causes more agreement errors than semantic gender because it is only syntactically specified, while semantic gender is both conceptually and syntactically specified. It is worth mentioning Vigliocco et al. (1999) could not find any differences between animate epicenes (referring to humans and animals) and inanimate epicenes (referring to objects). Thus, for Vigliocco et al. (1999), epicenes that refer to humans and animals seem not to be conceptually specified.

One could ask whether the results found by Vigliocco et al. (1999) and Vigliocco et al. (2005) contradict each other. The fact is that Vigliocco et al. (2005) do not mention Vigliocco et al. (1999). It may be hard to compare these two works since they report completely different goals, hypotheses, and tasks. In addition, the fact that epicenes referring to animals are conceptually motivated does not mean necessarily they are conceptually specified by sex as nouns with semantic gender are. Moreover, since epicenes that refer to animals are [+animate] they might be conceptually different from epicenes that refer to objects, which are [-animate], but this does not make them syntactically different from each other.

Moreover, the nature of stereotypical gender is still a debate on the literature. Oakhill et al. (2005) defended gender-stereotyped information is automatic and cannot be suppressed; however, Duff & Keir (2004) and Kneiner et al. (2008) showed stereotypical gender is automatic, but it can be suppressed by discourse disambiguation, or in Carreiras et al. (1996) by surface cues displayed in Spanish articles. On the other hand, Osterhout et al. (1997), Kennison & Trofe (2003), Cacciari & Padovani (2007) defended the idea stereotypical gender may be specified in the word's representation just like definitional gender; however, Kneiner et al. (2008), Canal et al. (2015) argue stereotypical gender and definitional gender are qualitative different. Stereotypical gender is pragmatically inferred by world knowledge probabilities as Oakhill et al. (2008) and Carreiras et al. (1996) affirmed, whereas definitional gender is lexically determined.

Finally, Casado et al. (2017), Kennison & Trofe (2003), Osterhout et al. (1997), and Canal et al. (2015) showed non-linguistic factors such as individual experiences and personal beliefs related to the sex of the participants could also influence gender retrieval. Casado et al. (2017) found out the sex of the

comprehenders can influence the way animate nouns are gender retrieved, including those nouns with grammatical gender. According to the authors, people encode and organize information according to their-selves sex role. Kennison & Trofe (2003) reported male comprehenders tend to rate stereotypical nouns more masculine, while female comprehenders tend to rate them more feminine. Osterhout et al. (1997) detected female comprehenders are either better at spotting gender mismatches or respond more strongly to violations of social expectations compared to male participants. Canal et al. (2015) claimed less feminine or less sexist comprehenders were less sensitive to gender mismatches.

3.12 Chapter digest

In this chapter, it was shown a gender could reveal interesting aspects regarding the lexicon organization and how morphemes are important in grammar. It was also curious to see how human mind can use language in different ways to refer to sexed beings and objects around us.

It was illustrated how the relation between morphemes is complicated since not all words that seem to share a root are actually a case of derivation (for example, case-casualty). Besides that, not everything that seems an affix is an affix indeed (as in submit and admit, for example). Morphology seemed so complicated and arbitrary that the generative grammar preferred to forget it, confining it to the lexicon (Lexicalist Hypothesis) or to the syntax (Distributed Morphology). Studies on morphology are becoming scarce, which is a shame since morphology is vital for some languages in the world.

In the group of Romance languages, Portuguese seems to be closer to Italian and Spanish in the terms of gender morphology. Gender is an intrinsic property of all nouns in Portuguese. Although all nouns in Portuguese have gender, not all nouns vary in gender. All animate nouns vary in gender with the purpose of identifying the sex of the referent. It is said these nouns have semantic gender. However, inanimate nouns and a few [+ human] nouns do not vary in gender. Their referents are not sexed marked. This class of nouns is called epicenes, and it is said they have grammatical gender. Moreover, there are plenty of double-gender nouns, also called common nouns, or bigenders. These nouns can refer to either sex without changing its form and meaning.

(78) Examples of nouns with semantic gender:

a. Masculine nouns: *professor* (male teacher), *engenheiro* (male engineer), *homem* (man);

b. Feminine nouns: *professora* (female teacher), *engenheira* (female engineer), *mulher* (woman).

c. Bigender nouns: *[o/a] estudante* (student), *[o/a]* recepcionista (receptionist), *[o/a]* surfista (surfist)

(79) Examples of epicenes in Portuguese:

a. Inanimate and abstract nouns: *[a] mesa* (table), *[a] fé* (faith), *[o] computador* (computer), *[o] amor* (love);

b. Animal names: *[a] tartaruga* (turtle), *[a] baleia* (whale), *[o] pássaro* (bird), *[o] gavião* (hawk).;

c. Nouns that refer to humans: *[a] vítima* (victim), *[a] pessoa* (person), *[o] indivíduo* (individual), *[o] gênio* (genius).

Portuguese has an overt redundant morphology so that articles, adjectives and pronouns must agree with nouns in gender. Moreover, Portuguese is a language with a phonological gender system, in which most nouns have transparent gender marks (feminine words end in –a and masculine words end in –o) and few words have opaque endings as *–e* as in *leite* (milk), *-l* as in *sol* (sun), *-n* as in *nuvem* (cloud), *-m* as in *imagem* (image), *-z* as in (*luz*), *-s* as in *dois* (two), *-r* as in *mar* (sea), etc. Furthermore, some words have irregular gender, that is, masculine words ending in *–a*, and feminine words ending in *–o* as in *problema* (problem) and *mão* (hand). Masculine gender is unmarked in Portuguese, and it also considered default.

For Villalva (1994), gender is not a formal category in Portuguese; it is a morphological property. She supported her claim by presenting examples in Portuguese, in which gender varies using lexical, derivational, and compositional processes.

It should be mentioned the results of psycholinguistic works on gender suggest gender morphology is crucial for lexical retrieval in Italian, Spanish and Portuguese so that nouns with different gender surface endings are processed differently. Transparent morphology helps gender retrieval in comparison to opaque morphology in all studies reported. Moreover, masculine nouns do not seem to be accessed through the mental lexicon, once they are unmarked and may not have an entry there. The studies reviewed in this chapter have not come to a final conclusion, but each of them has something to contribute:

(80)

a. All nouns are fully accessed in the mental lexicon and morphological access would happen at a later stage for checking purposes only (Bates et al., 1995);

b. All nouns are fully accessed in the mental lexicon and morphology would play a role since the beginning (Caffarra et al., 2015);

c. All nouns are accessed through morphological parsing in the mental grammar and opaque nouns depend on the gender conveyed in the articles (Afonso et al., 2013);

d. All transparent nouns can be either fully retrieved in the mental lexicon or retrieved by form in parallel, while opaque nouns may be fully retrieved in the mental lexicon (Caffarra et al., 2014);

e. Feminine dominant nouns are fully accessed in the mental lexicon, while feminine non-dominant and masculine nouns may be accessed by form (Corrêa et al., 2004);

f. All nouns are lexically retrieved in the mental lexicon according to frequency (Resende, 2015).

With respect to the types of gender, it seems grammatical gender is not so arbitrary it was supposed to be; epicenes that refer to humans can be influenced by meanings related to the sex of referents (Vigliocco et al., 2005) and the context bias (Corrêa et al., 1998). When compared to semantic gender, grammatical gender is not so easy to retrieve as semantic gender. What justifies semantic gender facilitatory effect is its intrinsic redundancy once it carries both conceptual and syntactic gender (Vigliocco et al., 1999).

Semantic gender can be expressed by whole-forms (definitional gender) such as *homem* (man) and *mulher* (woman) or compositional/derivational forms such as *engenheiro* (male engineer) and *engenheira* (female engineer). Definitional gender might be fully retrieved in the mental lexicon, while compositional / derivational gender might be retrieved via mental grammar, requiring more costly steps such as Affix Stripping. Thus, compositional/derivational gender is supposed to be more processing demanding than the definitional gender.

Semantic gender can also be ambiguous in the language, as for example, in bigenders such as *estudante* (student). Studies on pronoun processing have shown facilitatory effects for coreference with epicene antecedents in comparison to bigender attractors. The reason for a more costly processing of bigenders lies in the fact they depend on the context to retrieve their gender information (Cacciari et al., 2011; Alves, 2014).

Some bigenders, like role terms and occupations, can be stereotypically biased towards masculine as *surfista* (surfist) or feminine as *recepcionista* (receptionist) as in (80c). Stereotypes interference is still debatable in the literature, that is, some psycholinguists have shown evidence in favor of an automatic and unsuppressed influence of stereotypes, while others have shown it can be suppressed by context. Another discussion in the literature is regarding the comparison between stereotypical gender and definitional gender. Some studies have shown they seem to be both specified in the lexicon, while others have shown it is a result of a probabilistic-based account based on world knowledge inferences.

Finally, it seems non-linguistic factors related to the sex of comprehenders can affect gender retrieval and interpretation (Osterhout et al., 1997; Kennison & Trofe, 2003; Alves, 2014; Canal et al., 2015; Casado et al., 2017).

To conclude, more experimental evidence comparing different types of gender is needed. Some questions are still remaining: (a) are epicenes that refer to humans semantically motivated (Vigliocco et al., 2005)? To which extent are they different from nouns with semantic gender?; (b) is compositional semantic gender more processing demanding than semantic definitional gender? (Allen et al., 2003); (c) is stereotypical gender similar to definitional gender? (Kreiner et al., 2008; Osterhout et al., 1997; Canal et al., 2015); (d) does redundancy of surface cues facilitate processing? (Cacciari et al., 1997); (e) is masculine the default gender in two-gender Romance languages, that is, does masculine gender evoke both masculine and feminine representations? (Casado et al., 2017); and (f) does the sex of the comprehenders always affect gender retrieval (Osterhout et al., 1997; Kennison & Trofe, 2003; Casado et al., 2017)?

4. The present study

4.1 Outline

This chapter will focus on reporting the pre-Tests and the experiments conducted in this dissertation. In the first section, “Introduction”, some theoretical background previously discussed in the first two chapters that is important to the present study will be reviewed. In the second section, “Pre-Tests”, the pre-tests conducted for this dissertation will be reported. In the third section, “Experiments”, four eye-tracking experiments will be reported. In the fourth section, “General Discussion”, the main results of the pre-tests and the experiments will be summed up and compared.

4.2 Introduction

In chapter 1, pronoun binding theory was presented in the light of the Standard Binding Theory proposed by Chomsky (1981, 1986, 1993) and two Predicate-based Binding theories: Pollard and Sag (1992) and Reinhart & Reuland (1993). Principle B of Standard Binding Theory posits pronouns must be free in their governing category. On the other hand, Pollard & Sag (1992) claimed every pronoun must be co-indexed with none of its local obliqueness commanders (o-commanders), that is, it must not be co-indexed with a valent that precede the pronominal valence in the obliqueness order (subject > object > second object > PP > verb and predicate complements). Finally, Reinhart & Reuland (1993) excluded pronouns from binding theory as they were considered to be R-expressions, that is, independent referring expressions, as full NPs.

It should be highlighted though, that pronoun binding is not only governed by formal and structural rules. Kuno (1987) defended anaphoric binding domains should be expanded in order to include pragmatic and discourse factors such as discourse perceptiveness, definiteness, and the flow of discursive information (such topichood, comment and presupposition, and new and given information). Important pronoun discourse theories such as the Discursive Prominence Theory (Gordon & Hendrick, 1998) and the Centrality Theory (Grosz et al., 1995) must have inspired his ideas. According to Discursive Prominence Theory, pronouns must refer to high prominent entities, which are NPs located in high position in the syntactic tree. In addition, the semantic plausibility of the event described in the context should also be taken into account. For the Centrality Theory, coreference is related to coherence. This way, entities that are central, that is, in attentional states, are preferable pronoun referents. Grosz et al. (1995) suggested the attentional state also varies according to the grammatical role of the entity in discourse (subject > object > other).

Kehler and Rhode (2013) proposed a probabilistic model in which pronouns are equally influenced the world knowledge inferences and expectations (Hobbs, 1979) and discursive topichood (Grosz et al., 1995). This way, both top-down and bottom-up information would be simultaneously integrated in pronoun resolution. Kehler (2007) showed how the SMASH (Search, Match, and Select based on Heuristics) theory for pronoun resolution is deficient. He claimed the SMASH steps are not linear and independently separated as the model proposed.

With respect to the relationship between gender and discourse facts, Arnold et al. (2000) also argued in favor of a probability dynamic model of language processing in which both formal (gender cues) and discursive factors (accessibility and topichood) would be equally been taken into account. Rigalleau et al. (2004) claimed

pronouns do not create discursive accessibility (focus of attention), but they rather confirm it. The authors found out gender cues have an automatic influence in pronoun resolution, especially in languages as French. They explained in languages with grammatical gender, gender co-indexation must purely involve morphosyntactic features, and not conceptual gender regarding the sex of the referents. Foraker & McElree (2007) found out gendered pronouns facilitate resolution. They also claimed discourse accessible referents (located in the focus of attention) do not have faster pronoun retrievals, but they have an increased chance of being retrieved. This way, accessibility does not affect the retrieval cues, but it rather affects how the referent was encoded in memory.

When it comes to memory, the unitary-store memory model seems to be more appropriate to explain the efficiency and speed in which previous encoded information, called passive memory, is retrieved and integrated with new material, returning to the active memory again. One of the most relevant unitary-store memory models is the ACT-R content-addressable memory model. Lewis & Vasishth (2005) urged the items would be retrieved in memory in parallel according to their content cues. They predicted items that share content cues with the target would cause similarity-interference effects, that is, the strength of activation would be reduced due to the number of potential candidates, causing slower retrievals or even misretrievals. Since numerous results reported in the literature were not fully predicted by Lewis & Vasishth (2005)'s model. Engelmann et al. (2015) developed an extension to the content-addressable memory model, including distractor prominence [grammatical and discursive] e cue confusion (competition between same and different features) factors in the model.

The role of binding structural cues [Principle A and Principle B] and morphological cues [gender] was the object of study of various psycholinguists. Despite Engelmann et al. (2015) efforts to resolve all the contradictory results reported in the literature, there are at least two crucial questions that were addressable in none of those works: (a) whether languages with redundant visible morphology of gender would give more importance to gender agreement cues rather than structural constraints of Principle B; and (b) whether there is any difference regarding different gender cues.

Lago (2014) compared agreement attraction in subject-verb dependencies in Spanish (a language with redundant visible morphology) and English. Their results showed Spanish comprehenders showed more processing difficulties in ungrammatical sentences than English comprehenders. Moreover, Spanish comprehenders, but not English comprehenders, showed processing difficulties in grammatical sentences with plural distractors[i]. The authors explain that since agreement morphology is functionally more important in Spanish than in English, Spanish speakers would rely more on morphological cues in processing sentences. Therefore, the strength of agreement predictions would be higher for Spanish than in English, which causes a higher pay off when the predictions are not fulfilled and reanalysis is needed.

Taking into consideration the fact that the use of agreement cues may be more fruitful in languages with visible morphology, such as the Romance languages, the present work aims to investigate how pronouns retrieve antecedents in Brazilian Portuguese, which is also a language with visible morphology.

Moreover, it seems that the use of morphological cues in memory retrieval may also vary depending on the particular binding dependency. Agreement features

may be more helpful in pronominal antecedent retrieval due to the looseness of Principle B, since according to the Standard Binding Theory (Chomsky 1981, 1986, 1993), it only posits the pronoun antecedents must not be local.

> The recognition of a pronoun must initiate a retrospective search for an antecedent. Since the structural relation between a pronoun and its antecedent is almost free, it is natural do assume that a pronoun initiates a cue-based search for an antecedent that shares its person, number, and gender features, and hence it would not be surprising for this search to detect nouns that match those cues, even when they violate Principle B (PHILLIPS ET AL., 2011, 71)

This way, the present research will fill a gap in the literature, as it will investigate the role of morphological agreement in pronoun resolution in Portuguese, a language with redundant visible gender morphology. Furthermore, it will determine whether different types of gender features conveyed by antecedent candidates are responsible for differences in the way memory retrieves the antecedents, which was also never tested before.

In chapter 2, some pieces of experimental evidences have shown nouns could retrieve in different ways depending on their gender surface cues. A dual-route mechanism predicts regular forms (feminine nouns ending in –a and masculine nouns ending in –o) need to be decomposed by mental grammar, while irregular forms are fully accessed in the mental lexicon [Caffarra et al., 2014; Pinker, 1991; Allen et al., 2003]. Moreover, it seems that irregular forms might also depend on the context (definite articles) to retrieve gender information. On the other hand, in Portuguese, the evidences are still contradictory. Corrêa et al. (2004) found out feminine frequency dominant nouns are fully accessed in the mental lexicon, while masculine nouns would be morphologically parsed. On the other hand, Resende (2015) claimed in favor of a single-route mechanism, in which all gendered words would be retrieved in the mental lexicon through memory association.

Chapter 2 also reviewed experimental studies on different types of gender. Despite grammatical gender being traditionally treated as a purely arbitrary and formal gender [Corbett, 1991; Vigliocco & Franck, 1999; Rigalleau et al., 2004], Vigliocco et al. (2005) showed nouns with grammatical gender that refer to sexed beings as can be affected by conceptual gender.

It was also showed gender variation in nouns with semantic gender is expressed through composition/decomposition processes such as *menino* (boy) x *menina* (girl) and definitional gender such as *homem* (man) x *mulher* (woman). Moreover, bigenders, that is, nouns with ambiguous gender that are dependent on the context, can be gender stereotyped towards masculine or feminine as *surfista* (surfist) and *recepcionista* (receptionist).

Vigliocco & Franck (1999) compared nouns with grammatical gender (also called epicenes) with nouns with semantic gender. They found out agreement errors were more frequent in sentences with grammatical gender. They explained the facilitatory effect of semantic gender is a result of redundant information since semantic gender is both syntactically and conceptually specified.

Cacciari et al. (2011) and Alves (2014) compared epicene and bigender antecedents in pronoun resolution. They argued pronoun processing is faster with epicene antecedents than with bigender antecedents because the former has a fixed gender and syntactic in the language, while the latter has variable gender dependent on the context. However, Corrêa (2001) showed even pronoun resolution with epicene antecedents could take advantage of a gender-matching context.

The nature of stereotypical gender is still debatable in the literature. Some argue in favor of the idea it is specified in the lexicon just like definitional gender [Osterhout et al., 1997; Kennison & Trofe, 2003; Cacciari & Padovani, 2007], while

others defend stereotypical gender is qualitatively different from definitional gender [Kneiner et al., 2008; Canal et al., 2015], it might be pragmatically inferred by world knowledge probabilities [Carreiras et al., 1996; Oakhill, 2008].

The main aim of the present dissertation was to investigate how pronouns retrieve their antecedents in memory. Given the fact Portuguese is a language with visible morphology, and that pronouns seem to rely more on morphology in antecedent retrieval than, for example, reflexives, it was hypothesized that gender cues would play a great role in pronominal antecedent retrieval in Portuguese. It was expected to find effects of gender cues in both early and late antecedent retrieval processing phases. Moreover, it was expected structurally unacceptable candidates would be considered as potential candidates, influencing antecedent retrieval despite the fact they violate Principle B binding constraints.

In addition, it is hypothesized not only memory influences language, but also language influences memory. In other words, memory can be sensitive to different linguistic features; therefore, different gender types would be encoded/retrieved in memory with different weights. Although, van Dyke & McElree (2011) and Dillon et al. (2013) referred to cue weighting across different types of cues and grammatical dependencies. This dissertation will apply the concept of cue weighting across different types of genders. Thus the secondary aim of this dissertation is to determine which types of gender is more preferably retrieved in memory, that is, which types of gender is more weighed in memory. Consequently, effects of different types of gender were expected throughout antecedent retrieval processing. Antecedent candidates with compositional/derivational semantic gender were expected to be more preferably retrieved than epicene candidates, once the former are conceptually based. Antecedent candidates with definitional gender were expected to be more preferably retrieved

than candidates with stereotypical gender, as the former is lexically specified, while the former is resulted from a more complex process, a probabilistic model based on world knowledge inferences. On the other hand, antecedent candidates with definitional gender would be more preferably retrieved than compositional/derivational candidates once the former is fully retrieved in memory, while the latter would need to be processed by mental grammar. Finally, epicene candidates are expected to be more preferably retrieved than stereotypical gender.

In order to test the hypotheses, eye-tracking experiments were conducted with native speakers of Brazilian Portuguese. The eye-tracking technique is suitable for our purposes as it enables the researcher to examine the temporal course of language processing, including early (such as First Fixation Duration and First Pass) and late (such as Regression Path and Second Pass) on-line processing measures.

However, before conducting the eye-tracking experiments, pre-tests were carried out in order to appropriately select the materials to be used as antecedent candidates. Moreover, the pre-tests would elucidate the role of gender retrieval in word processing in Brazilian Portuguese and the differences between grammatical gender and stereotypical gender, a comparison never done before in the literature.

4.3 Pre-tests

As seen in the previous chapter, grammatical gender is invariable in the language and it must be specified in the word's representation [cf. Corbett, 1991; Carreiras et al., 1996; and Cacciari et al., 1997]. On the other hand, stereotypical gender can be either specified in the word's representation just like grammatical gender [cf. Osterhout et al., 1997; Kennison & Trofe, 2003; Cacciari & Padovani, 2007], or it is

probabilistically inferred with the help of world knowledge [cf. Oakhill et al., 2005; Carreiras et al., 1996; Kneiner et al., 2008; Canal et al., 2015].

The main purpose of the Pre-tests was to compare grammatical gender and stereotypical gender. The second purpose of the Pre-tests was selecting the most adequate gendered nouns to be used in the upcoming experiments of this dissertation.

Although, the intuition of the researcher and the lists of nouns contained in grammar books worked as a first-step in selecting the nouns, testing was necessary to guarantee the credibility of the experiments. It was thought comprehenders could interpret the gender of certain nouns differently from expected due to differences that may exist between the researcher's dialect and the dialect of the population to be tested. In addition, it was detected some nouns collected from grammar books were not very frequent in everyday language, which could negatively bias the results. Furthermore, since gender is also related to social issues, it was thought the gender of some nouns could have changed through time.

Pre-test 1, which was a cloze task, was used to determine the gender of the nouns to be used in the experiments. On the other hand, Pre-test 2, which as a Likert scale task, was used to detect the stereotypical gender of the same nouns. Therefore, the Pre-tests would cater for the gender of each noun (feminine, masculine, or ambiguous) and its stereotypical conceptual reference in the world.

4.3.1 Pre-test 1: agreement cloze task

Pre-test 1 was a gender assignment task inspired by Bates et al. (1996) and Resende (2015). However, unlike Bates et al. (1996) and Resende (2015), the interest was not only verifying the gender for feminine and masculine nouns, but also the gender of

ambiguous nouns. Thus a cloze agreement task was best suitable for this purpose so that gender would be assigned through agreement between definite articles and nouns.

The primary goal of Pre-test 1 was to verify gender assignment for epicenes referring to humans and stereotypical nouns. In other words, the objective of Pre-Test 1 was to check the syntactic gender of the nouns tested. Epicenes have grammatical gender, for example, *vítima* (victim) and *indivíduo* (individual) are syntactically determined as feminine and masculine respectively, but they can equally refer to both male and female referents. On the other hand, stereotypical nouns are bigender nouns (nouns with syntactic ambiguous gender and dependent on context) whose gender can be stereotypically biased such as *recepcionista* (receptionist) and *surfista* (surfist), which are feminine and masculine stereotyped respectively. The secondary goal was to select appropriate gendered nouns to be used in the eye-tracking experiments of this dissertation.

As any other epicenes, epicenes referring to humans have grammatical gender, and they are invariable in the language. Thus, it was expected female epicenes would be assigned as feminine (definite article *a*), and masculine epicenes would be assigned as masculine (definite article *o*), despite the fact they can both refer to either feminine or masculine referents. With respect to stereotypical nouns, it was expected they would be considered as ambiguous, receiving both feminine and masculine genders (definite article *a/o*). It should be noted since masculine gender is default in PB, masculine stereotypical nouns would be gender assigned more easily than feminine stereotypical nouns. Finally, for the same reason, it is predicted some neutral bigenders could be gender assigned as masculine more easily and more frequently than feminine.

If both grammatical and stereotypical genders are lexically specified in the word's representation, it is expected the participants will easily gender assign them both. On the other hand, if stereotypical gender is a result of a probabilistic inference based on world knowledge, while grammatical gender is formally specified in the lexicon, it is expected the participants will find more difficulties in assigning nouns with stereotypical nouns than nouns with grammatical gender. This prediction is based on the fact that lexical information seems to be more automatically retrieved than probabilistic information based on world knowledge inferences. In the latter case, calculating gender probabilities seems to be more psychologically demanding than retrieving information already specified in the lexicon.

With respect to the sex of the participants, it seems variables such as personal beliefs and the social status of females and males in a certain community seem to interfere with gender interpretation. Thus in order to test whether female and male participants would respond differently to gender assignment, the sex of the participants was also controlled. It was expected that female participants would gender assign feminine epicenes and feminine stereotypical nouns as feminine (definite article *a*) more frequently and more easily than male participants, while male participants are expected to gender assign masculine epicenes and stereotypical nouns as masculine (definite article *o*) more frequently and more easily than female participants [cf. Casado et al., 2017; Kennison & Trofe, 2003].

4.3.1.1 Participants

17 native speakers of Brazilian Portuguese (10 female and 7 male, with an average of age of 26 years) participated as volunteers in this experiment. They were recruited at

the Federal University of Rio de Janeiro (Rio de Janeiro, Brazil) and on social networks. All participants gave an informal consent to publish the results of this experiment. The undergraduate students received 3 hours of Cultural-Scientific Activities (*Atividades-Científico-Culturais Discentes*, AACC) as compensation for their work.

4.3.1.2 Materials and design

Our materials consisted of: a) 42 [possibly] neutral bigender nouns; b) 20 masculine stereotypical bigender nouns; c) 20 feminine stereotypical bigender nouns, d) 20 masculine epicenes; and e) 7 feminine epicenes[11]. The independent variables were: a) the type of the lexical item (bigender nouns or epicenes); b) the gender (masculine, feminine, or ambiguous); and c) the stereotype (masculine, feminine, or neutral). The dependent variables were the responses of the participants and their reaction times.

	Type of lexical item	Gender	Stereotype
Neutral bigender	Bigender	Ambiguous	Neutral
Masculine stereotypical bigender	Bigender	Ambiguous	Masculine
Feminine stereotypical bigender	Bigender	Ambiguous	Feminine
Masculine epicene	Epicene	Masculine	Neutral
Feminine epicene	Epicene	Feminine	Neutral

Table 2: Discriminating the conditions of Pre-test 1

[11] There were very few feminine nouns with grammatical gender in the materials due to the fact that there are not many of them referring to humans in BP.

One can find a sample of the materials used in the experiment in Table 3, while a complete list of the materials can be found in appendix.

Neutral bigender	adolescente (*teenager*)
Masculine stereotypical bigender	surfista (*surfist*)
Feminine stereotypical bigender	recepcionista (*receptionist*)
Masculine epicene	indivíduo (*individual*)
Feminine epicene	vítima (*victim*)

Table 3: Pre-tests materials sample

4.3.1.3 Procedure

The experiment was run on *Ibex Farm*[12] on-line platform, and the participants received a link to our experiment webpage. They filled a form with their basic personal information, received instructions for the experiment as showed in Figure 6, and practiced with 8 trials in order to get adapted to the task. They were instructed to perform the task in a quiet place without any distractions. Figure 6 shows the task screen the participants were exposed to.

[12] Drummond, A. (2018). Ibex Farm. URL http://spellout.net/ibexfarm

Instructions

You will read silently the words that will pop up on your screen. Then you will be asked to choose the more adequate definite article for that word.

For example:

"_____ turista"
"a"
or "o"
or "a/o"

If you think the most adequate answer is "a", you should choose the first option, if you think it is "o", you should choose the second option, but if you think "a" and "o" are equally possible, you should choose the third option, "a/o".

Don't feel worried about marking the right answers, choose the first answer that comes to your mind.

Please, answer as quickly as you can, in less than 1 second.

If you've understood the task and you're ready to start, please, avoid distractions (as your telephone) so that you can keep focused.

When ready, press the button bellow.

Click here to continue.

Figure 5: Pre-test 1 instructions

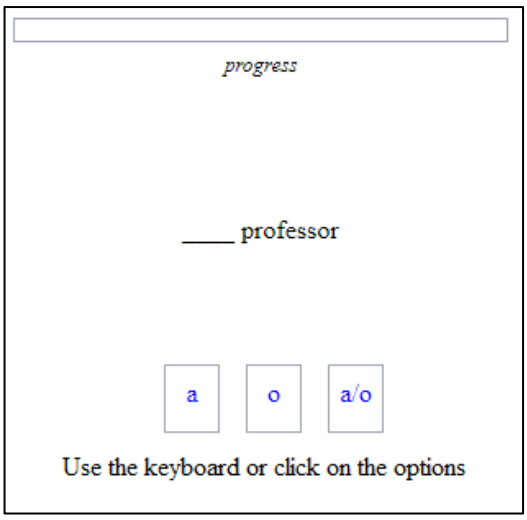

Figure 6: Pre-test 1 task screen

4.3.1.4 Analysis

The data for reaction times, that is, the time it took the participants to press the button to choose an answer, was analyzed using the software R[13]. Since the items chosen and the participants recruited for the task could cause noise to the results due to unknown factors, linear mixed effects models (*lmes*) were better suited for the present data. They were created with the *lmerTest* package[14]. This way, such variables could work as random effects, whereas the independent variables could work as fixed effects. This type of analysis "is superior to the traditional repeated-measures and mixed ANOVAS approaches" (LEVSHINA, 2015, p.193). Therefore *lmes* were created with random intercepts for items and participants (random slopes did not converge), which means individual adjustments would be made for the intercept for each individual and

[13] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (https://www.r-project.org/)

[14] Alexandra Kuznetsova, Per Bruun Brockhoff and Rune Haubo Bojesen Christensen (2015). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-29 (https://CRAN.R-project.org/package=lmerTest)

item. After the models were created, the *anova ( )* function was applied to find out which contrasts were statistically relevant.

The first *lme* was composed by the variables *condition* and *sex of the participants* as fixed effects and *items* and *participants* variables as random effects. The second model was composed by the variables *type of lexical item* (bigender or epicene), *gender* (masculine or feminine), *stereotype* (masculine biased, feminine biased, or neuter), and *sex of the participants* as fixed effects, while *items* and *participants* were considered random effects[15].

On the other hand, the answers of the participants data were analyzed using the Chi-squared test (χ^2-test) of independence to test whether the difference between the conditions were statistically significant. Two χ^2-tests were conducted to test whether the difference among the answers were associated to either the conditions or the sex of the participants. The frequency of each type of answer was reported according to each condition and the sex of the participants in percentage.

4.3.1.5 Results

Reaction times

It is important to mention we did not find any statistically significant differences for reaction times in neither of our models. The means of the reaction times with the standard deviations in parenthesis for each condition according to the sex of the participants can be found in Table 4.

[15] There was an attempt to formulate a model with random slopes having *sex of the participants* as a random coefficient of the *participants* effect; however, this model did not converge.

Conditions	Male Participants	Female Participants	Total
Neutral bigender	2513 (1517)	2922 (5430)	11821 (147621)
Masculine stereotypical bigender	2407 (1420)	2831 (2676)	2957 (3379)
Feminine stereotypical bigender	2584 (2009)	3258 (6071)	2969 (4748)
Masculine epicene	32045 (277141)	21714 (21250)	24859 (235456)
Feminine epicene	3069 (2332)	2603 (1632)	2803 (1966)

Table 4: Pre-test 1 reaction times in milliseconds and standard deviations in parenthesis

Cloze Responses

The results of the χ^2 tests revealed the there was a significant association between the participants' responses and the conditions: $\chi^2(8) = 788.87$, $p<0.05$. In addition, there was as a trend towards significance between the participants' responses and their sex: $\chi^2(2) = 5.394$, $p=0.067$.

The participants' responses were reported in Tables 5 and 6. Table 5 reports their responses in percentage by row, while Table 6 by column.

The following nouns exemplify each condition as the following:

(81)

i) Neutral bigender noun: *turista* (tourist)

ii) Masculine stereotypical noun: *surfista* (surfist)

iii) Feminine stereotypical noun: *recepcionista* (receptionist)

iv) Masculine epicene noun: *indivíduo* (individual)

v) Feminine epicene noun: *vítima* (victim)

	Neutral bigender	Masculine stereotypical bigender	Feminine stereotypical bigender	Masculine epicene	Feminine epicene
A	21.7	3.6	34.8	1.8	38.0
a/o	51.1	20.2	18.7	8.2	1.5
O	26.3	22.8	5.2	45.2	0.2

Table 5: Pre-test 1 answers of the participants in percentage by row

	Neutral bigender	Masculine stereotypical bigender	Feminine stereotypical bigender	Masculine epicene	Feminine epicene
A	8.3	3.0	30.5	1.5	85.7
a/o	73.3	62.7	61.1	26.9	13.2
O	18.2	34.2	8.3	71.4	1.0

Table 6: Pre-test 1 answers of the participants in percentage by column

The answers according to the sex of the participants were reported in Figure 7.

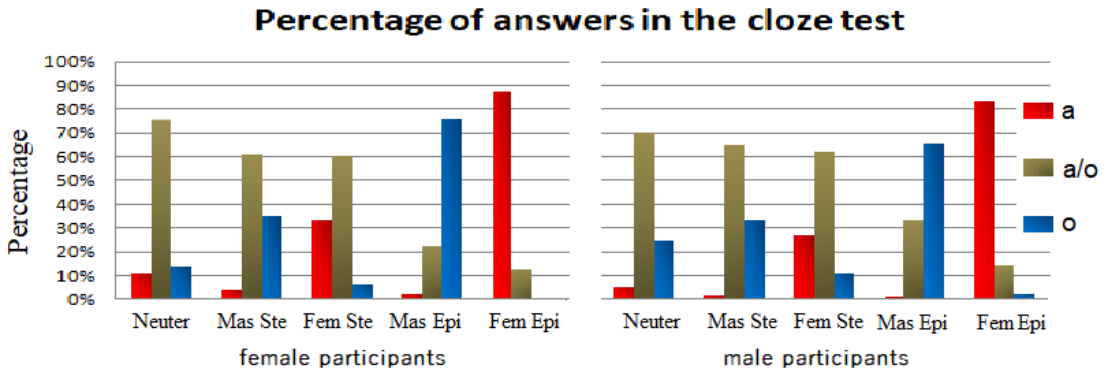

Figure 7: Pre-test 1 answers by sex of the participants

4.3.1.6 Discussion

The reaction times (*on-line* data) of the of Pre-test 1 did not show any difference between epicenes and stereotypical nouns, which might indicate those two types of nouns retrieve gender information similarly. Thus, it seems both grammatical and

stereotypical gender information is lexically specified in the noun's representation [Osterhout et al., 1997; Kennison & Trofe, 2003; and Cacciari & Padovani, 2007].

With respect to the responses of the participants (*off-line* data), neutral bigender nouns received more "a/o" answers (51.1%) than the stereotypical bigenders (20.2% for masculine stereotypical nouns and 18.7% for feminine stereotypical nouns), as it was expected. Although neutral nouns were not supposed to present any gender stereotype, they were inclined to be rather masculine gender assigned (18.2%) than feminine gender assigned (8.3%). This result was not a surprise since masculine is the default gender in Portuguese. Some neutral bigenders that were masculine gender assigned were: *banhista* (bather), *hóspede* (guest), *informante* (informant), *repórter* (reporter), *romancista* (romanticist), *sem-teto* (homeless person), etc.

Unexpectedly, there was no difference in the responses according to the type of stereotype. In contrast to what was predicted, the gender of the epicenes seems to matter, as feminine epicenes received more "a" responses (85.7%) than masculine epicenes received "o" responses (71.4%). It seems feminine epicenes are more marked than masculine epicenes.

It is worth noting it was expected some epicenes would not behave like true epicenes, that is, some epicenes would accept ambiguous gender assigning; however, there was an expressive amount of "a/o" responses assigned to epicenes (26.9% for masculine epicenes and 13.2% for feminine epicenes). This could be have been caused by dialect peculiarities, or a possible language change. Since epicenes that refer to humans are very limited in number in BP, they are showing a tendency to behave like bigender nouns. What epicene and bigender nouns have in common is the fact they can refer to either female or male referents. In this case, masculine epicenes received more "a/o" than feminine epicenes because masculine is the default gender

in BP, that is, masculine gender naturally tend to refer to both female and male referents. Some masculine epicenes that were gender assigned as bigenders were the following: *bebê* (baby), *cônjuge* (spouse), *dedo duro* (snitch), *neném* (baby), etc.

Although the factor *sex of the participants* only showed a trend towards significance, it seems male participants assigned more "a/o" answers to masculine epicenes than female participants, which could contribute to the idea that female speakers seem to be more conservative to language violations [Osterhout et al., 1997].

4.3.2 Pre-test 2: Gender bias judgment

Pre-test 2 was a rating study, whose primary goal was to examine how epicenes and stereotypical nouns conceptually process their referents, while the secondary goal was to check gender stereotypes in BP in order to select which stereotypical nouns would be used in the upcoming experiments. This step is mandatory for any research whose object of study is gender stereotypes.

It was predicted stereotypical nouns, but not epicenes, would be stereotypically rated since epicenes can equally refer to either male or female referents (Corbett, 1991; Cacciari et al., 1997). Moreover, it is expected neutral bigenders would be rated as more masculine stereotyped than feminine stereotyped once masculine gender is the default gender in BP. For the same reason, masculine stereotypical nouns would be rated slower than feminine stereotypical nouns.

With respect to the sex of the participants, male participants were expected to rate nouns more masculine biased than female participants, similarly, female participants were expected to rate nouns more feminine biased than male participants [Casado et al., 2017; Kennison & Trofe, 2003].

4.3.2.1 Participants

17 native speakers of Brazilian Portuguese (10 female and 7 male, with an average of age of 24 years) participated as volunteers in this experiment. They were also recruited at the Federal University of Rio de Janeiro (Rio de Janeiro, Brazil) and on social networks. All participants gave an informal consent to publish the results of this experiment. The undergraduate students received 3 hours of Cultural-Scientific Activities (*Atividades-Científico-Culturais Discentes*, AACC) as compensation for their work.

4.3.2.2 Materials and design

Our materials were the same of Pre-test 1.

4.3.2.3 Procedure

The procedure[16] was the same of Pre-test 1, except for the task. In this experiment the participants were instructed to judge the nouns based on their gender stereotype in a Likert scale from 1 to 5 (1 for feminine and 5 for masculine). One can see the instructions in Figure 8 and task screen in Figure 9.

[16] It was very difficult to investigate gender bias as it is a taboo. One of my participants complained about the instructions and the task itself as it excluded the LGBTQ+ community. I apologize for that, but since the object of study did not involve issues on the diversity of genders and sexes that exist in society, it was preferred to keep the binary paradigm - feminine and masculine - instead.

Instructions
You will read silently the words that will pop up on your screen. Then you will be asked to use a scale to indicate the sex to which that word refers in your opinion.
For example:

"tourist"

(1) for very female
or (2) for female
or (3) for neuter
or (4) for male
or (5) for very male

If you think "tourist" is a word used almost exclusively to refer to people of the feminine sex, you should choose the first option; if you think "tourist" is a word generally used to refer people of the feminine sex, you should choose the second option; if you think "tourist" is a word used to refer to people of both feminine and masculine sexes, you should choose the third option; if you think "tourist" is a word used almost exclusively to refer to people of the masculine sex, you should choose the fourth option; if you think "tourist" is a word generally used to refer people of the masculine sex, you should choose the last option.
Don't feel worried about marking the right answer; choose the first answer that comes to your mind.
Please, answer as quickly as you can, in less than 1 second.
If you've understood the task and you're ready to start, please, avoid distractions (as your telephone) so that you can keep focused.
When ready, press the button bellow.
Click here to continue.

Figure 8: Pre-Test 2 instructions

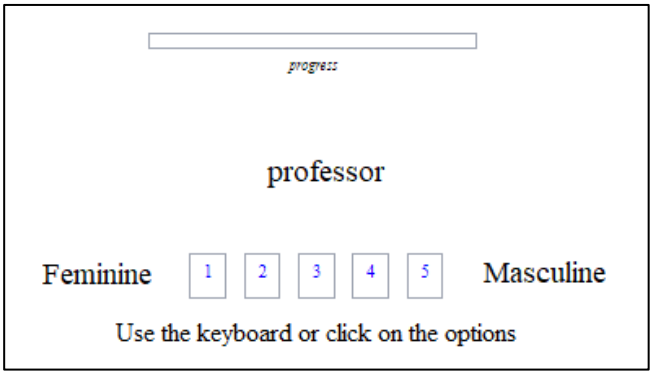

Figure 9: Pre-test 2 task screen

4.3.2.4 Analysis

The same of Pre-test 1; however, since the task involved a Likert scale, the answers of the participants were transformed into z-scores in order to minimize the individualities that the participants might have in relation to the scale. Thus each response of each participant was calculated based on the means and the standard deviations of the means for that single participant. Consequently, Pre-test 2 ended up with 2 continuous variables – the reaction times and the z-scores, which were analyzed using the same factors of the *lmes* of Pre-test 1.

After creating the *lmes*, ANOVAs were run in order to derive F statistics and p-values (using the Satterthwaite approximation for degrees of freedom). Consequently, the *lmes* created could be proved statistically significant or not. However, ANOVAs only tells there is some significance somewhere, but it cannot tell where. Therefore, parametric *Tukey Honest Significant Differences* (HSD) were conducted as post-hoc tests to check which conditions are statistically significant since it returns the 95% confidence intervals of the differences between the groups. This kind of test is quite robust to violations of normality and it avoids the inflation of surprise, which happens when there are too many conditions and pairwise comparisons (Levshina, 2015).

It is worth mention one token of the reaction times was excluded since it lasted more than 5 seconds.

4.3.2.5 Results

Reaction Times Results

The ANOVAs results of the first model indicated a significant main effect of conditions in the first model, $F(4, 1934633) = 2.55$, $p = 0.043$), and a marginally significant interaction between conditions and sex of the participants, $F(4, 1934633) = 2.26$, $p = 0.060$). Table 7 illustrates the means of the reaction times with standard deviations for each condition according to the sex of the participants.

Conditions	Male Participants	Female Participants	Total
Neutral bigenders	2213 (1691)	2198 (1551)	2204 (1455)
Masculine stereotypical bigenders	1879 (1689)	2246 (1549)	2103 (1173)
Feminine stereotypical bigenders	2568 (1689)	2319 (1549)	2415 (1875)
Masculine epicenes	2396 (1692)	2528 (1550)	2477 (1832)
Feminine epicenes	1877 (1563)	2127 (1552)	2415 (4446)

Table 7: Reaction times in milliseconds and standard deviations in parenthesis in Pre-test 2

Figure 10 illustrates the means of the reaction times by condition.

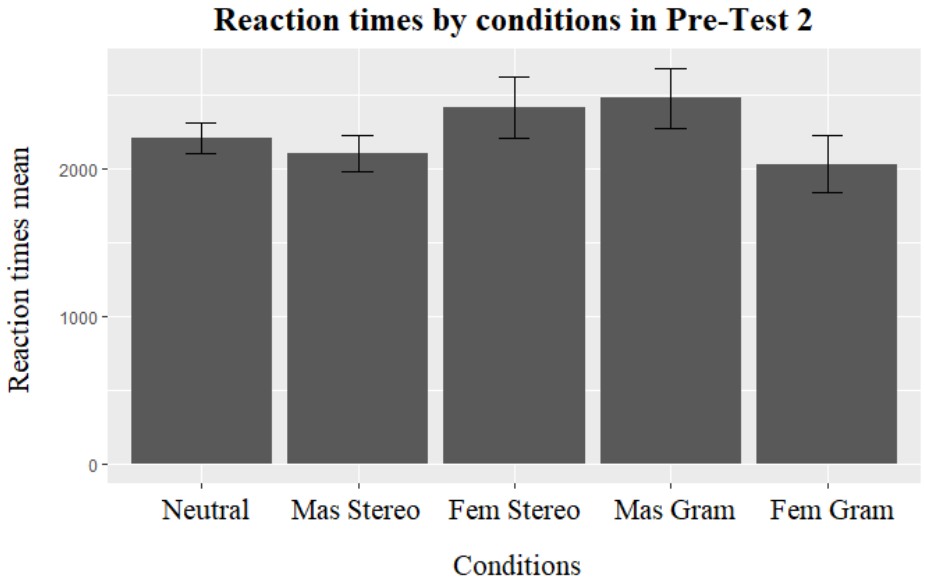

Figure 10: Reaction times by conditions in milliseconds in Pre-test 2

In order to check which pairs of conditions were statistically significant, parametric Tukey Honest Significant Differences (HSD) was conducted as post-hoc tests. The results indicated masculine epicenes had 374ms longer reaction times than masculine stereotypical bigenders (p=0.015) and 445ms longer than feminine epicenes (p=0.048). Although the results were marginally significant, it should be mentioned the reaction times for the neutral bigenders were 273ms longer when compared to masculine epicenes (p=0.060). The masculine stereotypical bigenders were 312ms faster than the feminine stereotypical ones (p=0.024*), especially for male participants as this difference increases up to 688ms (p=0.004**). It should be mentioned Tukey HSD tests did not find any statistically significant difference between the conditions according to the sex of the participants.

The interaction between *conditions* and *sex of the participants* can be visualized in the interaction plot in Figure 11.

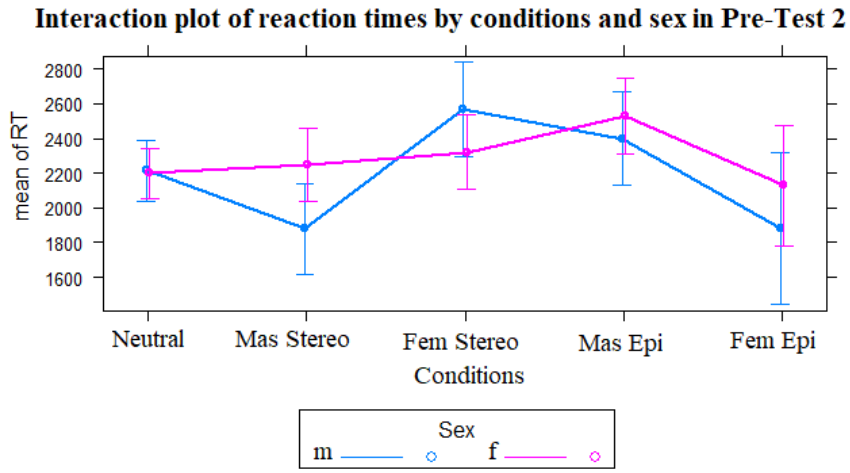

Figure 11: Interaction plot of reaction times in milliseconds by conditions and sex of the participants in Pre-test 2

As shown in the interaction plot below and according to the Tukey HSD tests, male participants rated nouns with masculine stereotypical gender 688ms faster than

they rated nouns with feminine stereotypical gender (p=0.012). Male participants also rated masculine stereotypical gender 517ms faster than they rated feminine epicenes in a trend towards significance (p=0.174).

In the second model, it was detected a significant main effect of *gender*, F(1, 1934633) = 4.08, p = 0.045, a significant main effect of *stereotype*, F(1, 1934633) = 5.14, p = 0.025), and a significant interaction between *stereotype* and *sex of the participants*, F(1, 193463) = 7.80, p = 0.005).

According to HSD Tukey Tests, in general, masculine nouns were rated 445ms longer than feminine nouns (p=0.016), which were rated 321ms faster than neutral bigenders (p=0.065), although in the latter case with a marginal statistically significance. In addition, masculine stereotypical nouns were rated 312ms faster than feminine stereotypical nouns (p=0.024), which replicated the results for the first model.

The interaction plot below was created with the purpose of examining the interaction between *stereotype* and *sex of the participants*.

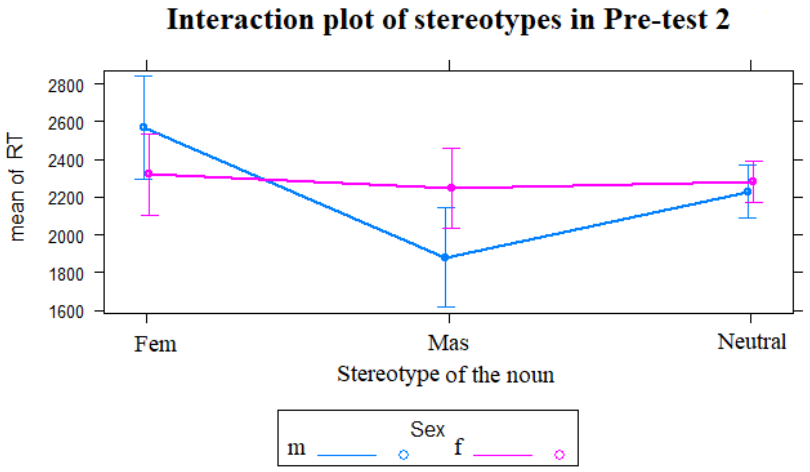

Figure 12: Interaction plot of reaction times by stereotypes and sex of the participants in Pre-test 2

According to Figure 12 and Tukey HSD tests, male participants were 688ms slower at rating nouns with feminine stereotype than with masculine stereotype (p=0.004).

Z-scores results

The ANOVAs results of the first model indicated a significant main effect of *conditions* in the first model, $F(4, 0.514) = 53.24$, $p < 0.05$), and a significant interaction between *conditions* and *sex of the participants*, $F(4, 0.514) = 3.65$, $p = 0.005$). The means of the Z-scores and standard deviations by condition and according to the sex of the participants were reported in Table 8.

Conditions	Male Participants	Female Participants	Total
Neutral bigender	-0.047 (0.925)	-0.119 (0.892)	-0.091 (0.700)
Masculine stereotypical bigender	0.438 (0.923)	0.315 (0.892)	0.363 (0.836)
Feminine stereotypical bigender	-0.885 (0.923)	-0.777 (0.892)	-0.819 (0.856)
Masculine epicene	0.337 (0.925)	0.363 (0.888)	0.353 (0.868)
Feminine epicene	-0.320 (0.864)	-0.710 (0.887)	-0.537 (0.782)

Table 8: Z-scores and standard deviations in parenthesis by conditions and according to the sex of the participants in Pre-test 2

Figure 13 illustrates the Z-scores by conditions.

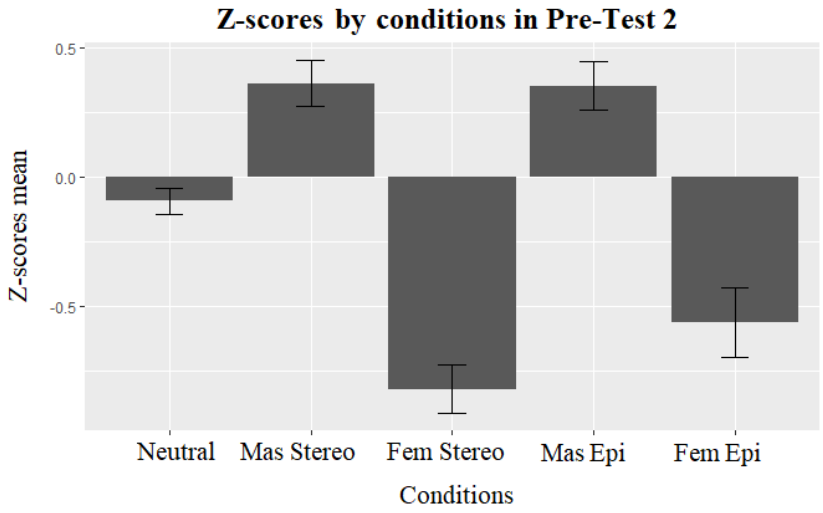

Figure 13: Z-scores by conditions in milliseconds in Pre-test 2

Tukey Tests HSD tests revealed significant differences between all conditions, except for masculine stereotypical bigenders and masculine epicenes. The neutral bigenders worked as a control condition as they were approximately rated in the middle of the scale. Taking the neutral bigenders as reference, the masculine stereotypical nouns and the masculine epicenes were more right-oriented in 0.45 ($p<0.05$) and 0.44 ($p<0.05$) z-scores respectively. On the opposite direction, the feminine stereotypical bigenders and the feminine epicenes were more left-oriented in -0.72 ($p<0.05$) and -0.44 ($p<0.05$) z-scores respectively. The feminine stereotypical bigenders were more left-oriented in -1.18 z-scores than the masculine stereotypical bigenders ($p<0.05$) while the feminine epicenes were more left-oriented in -0.89 z-scores than the masculine epicenes ($p<0.05$). Finally, the feminine stereotypical bigenders were more left-oriented in 0.28 z-scores than the feminine epicenes ($p=0.006$).

One can visualize the z-scores by conditions and sex of the participants in Figure 14:

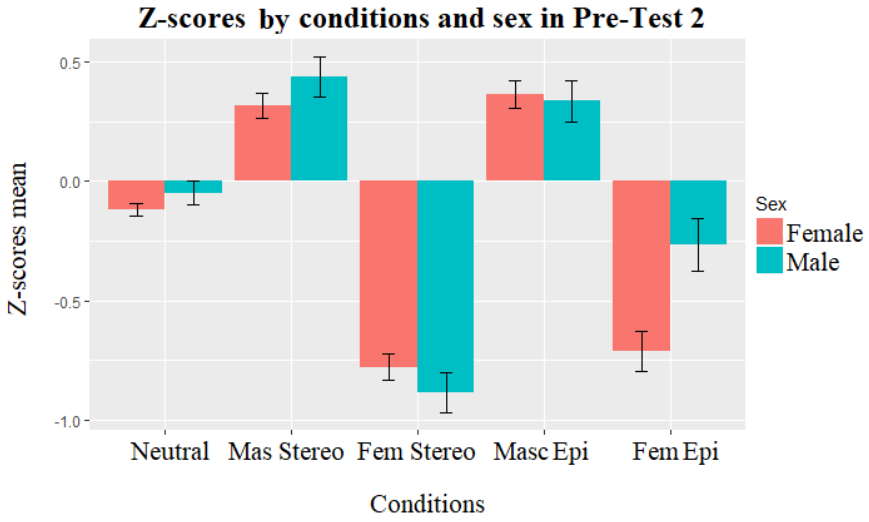

Figure 14: Z-scores by conditions and sex of the participants in Pre-test 2

When the sex of the participants was taken into account as Figure 12 above illustrates, Tukey HSD tests showed a marginal significance between male and female participants in rating feminine epicenes. Male participants rated them 0.44 z-scores more right-oriented than female participants ($p=0.062$). In addition, female and male participants treated the contrast between the conditions differently. Female participants rated the masculine epicenes more right-oriented than the feminine epicenes in 1.07 z-scores ($p<0.05$), while this difference dropped to 0.60 z-scores ($p<0.05$) for male participants. The contrast between the bigenders in general and the epicenes was also felt differently according to the sex of the participants. Female participants rated the masculine epicenes more right-oriented than the bigenders in 0.30 z-scores ($p<0.05$) and the feminine epicenes more left oriented than the bigenders in -0.77 z-scores than male participants, 0.16 z-scores ($p<0.05$) and -0.43 ($p<0.05$) respectively.

The ANOVAs results of the second model indicated significant main effects of *gender*, $F(1, 0.514) = 40.45$, $p < 0.05$; and *stereotype*, $F(1, 0.514) = 154.36$, $p < 0.05$; significant interactions between *gender* and *sex of the participants*, $F(1, 0.514)$

= 9.33, p = 0.002; and between *stereotype* and *sex of the participants*, F(1, 0.514) = 4.02, p <0.045.

In general, feminine nouns were more left-oriented rated than bigenders in -0.64 z-scores ($p<0.05$), while masculine nouns were more right-oriented rated than bigenders in 0.24 z-scores ($p<0.05$). Feminine nouns were more left-oriented rated than masculine nouns in 0.89 z-scores ($p<0.05$). Similarly, as a whole, feminine stereotypical bigenders were more right-oriented rated than neutral bigenders in 0.60 z-scores ($p<0.05$), while masculine stereotypical were more left-oriented than neutral bigenders in -0.57 z-scores ($p<0.05$). Masculine stereotypical bigenders were more right-oriented in 1.32 z-scores than feminine stereotypical bigenders ($p<0.05$).

Furthermore, male participants rated feminine nouns more right-oriented than female participants in 0.39 z-scores ($p=0.06$), even though it was a trend towards significance. Although, ANOVAs did not show any effects of type of lexical item and type of gender, Tukey HSD tests detected, in general, epicenes were more right-oriented rated than the bigenders in general in 0.25 z-scores ($p<0.05$). Finally, the socially determined genders (stereotypical bigenders) were more left-oriented rated than the linguistically determined genders (epicenes and neutral bigenders together) in -0.09 z-scores ($p=0.009$).

The results indicated a main effect of *gender* and *bias of the nouns* with interactions between each of those two factor and the *sex of the participants*. Figures 15 and 16 illustrates the nature of those interactions.

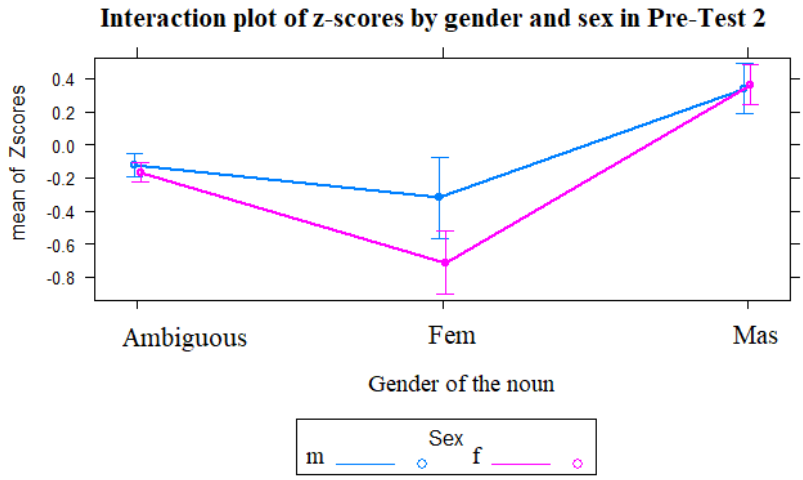

Figure 15: Interaction plot of z-scores by gender of the nouns and sex of the particiopants in Pre-test 2

In Figure 15 and according to Tukey HSD tests, male participants rate feminine nouns 0.39 z-scores more right-oriented than female participants in a trend towards significance (p=0.064).

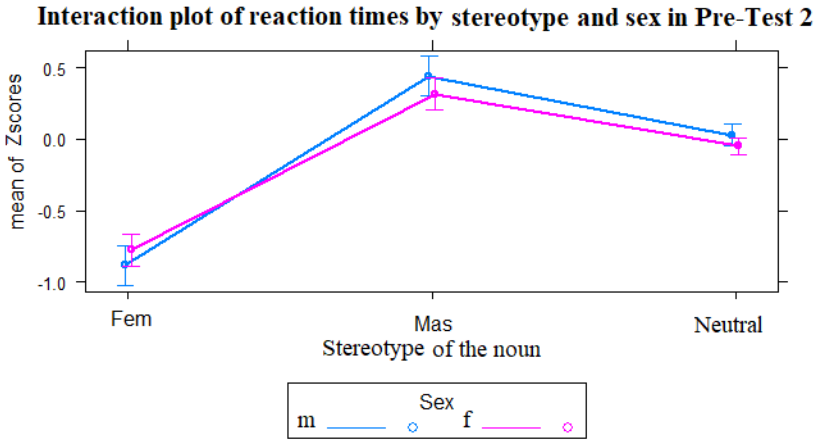

Figure 16: Interaction plot of reaction times by bias and sex of the participants in Pre-test 2

No difference regarding the sex of participants and noun stereotypes were found by Tukey HSD tests.

4.3.2.6 Discussion

Pre-test 2 investigated how speakers of BP treat gender bias in nouns. We tested five different conditions: (a) neutral bigenders, (b) masculine stereotypical bigenders, (c) feminine stereotypical bigenders, (d) masculine epicenes (nouns with masculine grammatical gender), and (e) feminine epicenes (nouns with feminine grammatical gender).

The *on-line* results indicated masculine stereotypical bigenders were conceptually processed faster than masculine epicenes. Masculine stereotypical bigenders might be underspecified for gender, requiring pragmatic inferences based on world knowledge probabilities to conceptually identify their referents as postulated by Oakhill et al. (2005), Carreiras et al. (1996), Kneiner et al. (2008), and Canal et al. (2015). On the other hand, masculine epicenes may carry two conceptual representations, a masculine and a feminine. This way, having underspecified conceptual gender was crucial for masculine stereotypical bigenders have faster conceptual processing than masculine epicenes.

By comparing masculine and feminine epicenes, the *on-line* results indicated the latter were conceptually processed faster than the former. One could argue the *–a* ending superficial cues of the feminine epicenes facilitated processing (Cacciari et al., 1997), because they match the semantic *gender to sex* rule (Vigliocco et al., 2005) as most feminine nouns end in *–a* in BP. However, most masculine epicenes end in *–o*, also matching the semantic *gender to sex* rule of masculine nouns in BP. Thus, superficial cues could not explain feminine epicenes facilitation. An alternative explanation is the following: despite not being numerous, feminine epicenes are more frequent than masculine epicenes; therefore, feminine epicenes might be easily

retrieved from the mental lexicon (Corrêa et al., 2004; Resende, 2015). But since this research did not control for frequency effects, it might be hard to support this idea[17] here. Another explanation, then, lies in the fact that feminine epicenes are gender-marked and must be more conceptually salient, suffering less influence from masculine default gender. This might explain the reason why feminine epicenes were conceptually processed faster than masculine epicenes.

While masculine epicenes were conceptually masculine biased probably because of the influence of masculine default gender, neutral bigenders were conceptually ambiguous because they cannot count on world knowledge probabilities to infer their gender. In other words, neutral bigenders are dependent on the linguistic context, and since there was no given context in Pre-test 2, neutral bigenders tended to be slowly processed when compared to masculine epicenes.

In agreement with the predictions, masculine stereotypical bigenders were conceptually processed faster than feminine stereotypical bigenders due to the fact world knowledge probabilities might be more easily processed for masculine than feminine. A linguistic reason for this is related to the influence of masculine default gender, and a non-linguistic reason is related to the fact that the materials used in task were mostly role terms, and as there are more male referents than female referents occupying work positions in society, it must be easier to calculate probabilities for masculine stereotypical nouns.

Contrary to predictions, in general, masculine nouns were more slowly rated than feminine nouns, which were rated faster than neuter nouns. In this case, as masculine gender is the default gender, masculine might convey two conceptual representations, a masculine and a feminine, while feminine might convey only one

[17] Frequency effects were not controlled in this research due to the limited number of epicenes referring to humans in BP.

conceptual representation, a feminine one. That is the reason why it was more difficult to conceptually process masculine nouns than feminine nouns (Corbett, 1991; Cacciari & Padovani, 2007; Casado et al., 2017). Similar to masculine nouns, neutral bigenders also convey two gender representations; however, they depend on context to decide whether it is the masculine or the feminine representation is going to be used, and since no context was given in the task, they were conceptually processed more slowly than feminine nouns.

With respect to *off-line* results, in general, epicenes were rated as more masculine than bigenders, which is evidence that epicenes are more influenced from masculine default gender than bigenders. Epicenes might convey two conceptual representations, while bigenders are underspecified and their conceptual representations may be inferred from context and world knowledge probabilities. Similarly, neutral bigenders were rated as more masculine than stereotypical bigenders. These results suggest world knowledge inferences do not suffer much influence from default gender as conceptual representations. World knowledge may be dependent on pragmatics, while conceptual representations may be dependent on semantics. Thus default gender seems to affect semantics more than pragmatics.

The differences between female and male participants were also very interesting. As expected, male comprehenders seem to be more easily sensitive to masculine gender than female comprehenders, which is in congruence with Kennison & Trofe (2003) and Casado et al. (2017).

4.3.3 Pre-tests general discussion

In Pre-test 1, it was not detected any on-line difference regarding how nouns with grammatical gender and stereotypical gender were gender assigned. However, in Pre-test 2, the on-line results indicated nouns with grammatical gender and stereotypical gender were processed at different ways. At first glance, these results sound contradicting, but one should keep in mind Pre-test 1 and Pre-test 2 required different types of processing. Pre-test 1 was a gender-assigning task, while Pre-test 2 was a gender stereotypical rating task. In Pre-test 1, participants were required to inform the syntactic gender of the nouns, while, in Pre-test 2, participants were required to inform the conceptual gender of the nouns. Thus, it seems the gender information of both grammatical gender and stereotypical gender are similarly stored, but their conceptual gender is processed differently.

Both grammatical gender [Corbett, 1991; Carreiras et al., 1996; and Cacciari et al., 1997] and stereotypical gender [cf. Osterhout et al., 1997; Kennison & Trofe, 2003; Cacciari & Padovani, 2007] might be stored in the word's representation. This means that nouns with these types of genders are probably fully retrieved in the mental lexicon. According to the results of Pre-test 1, the syntactic gender of masculine epicenes is masculine and the syntactic gender of feminine epicenes is feminine. On the other hand, the syntactic gender of stereotypical and neutral bigenders is underspecified. It seems that having underspecified gender does not guarantee bigenders and stereotypical bigenders faster gender assigning, which means that in syntactic terms, they are similar.

However, the differences between epicenes and bigenders only emerge when conceptually gender processing is analyzed. Masculine stereotypical bigenders were conceptually processed faster than masculine epicenes. Thus there are reasons to believe that stereotypical bigenders are also underspecified for conceptually gender,

while epicenes are ambiguous. In other words, they both admit a female and a male conceptual representation, but stereotypical bigenders do so faster. The fact is, although stereotypical bigenders have conceptually underspecified gender, they are capable of pragmatically infer their conceptual gender based on world knowledge probabilities [cf. Oakhill et al., 2005; Carreiras et al., 1996; Kneiner et al., 2008; Canal et al., 2015]. On the other hand, although it is not necessary, epicenes are also capable of semantically infer their conceptual gender (Vigliocco et al., 2005; Konishi, 1993). In other words, masculine stereotypical bigenders have underspecified conceptual gender combined with pragmatic probabilities, while masculine epicenes have ambiguous conceptual gender combined with semantic factors. From these two types of conceptual genders, masculine stereotypical bigenders are processed faster than masculine epicenes probably because the former have underspecified conceptual gender.

Like stereotypical bigenders, neutral bigenders may also have underspecified conceptual gender. But unlike them, neutral bigenders cannot count on probabilities; they require the help of linguistic context to have their gender inferred. If there is no context available, their conceptual gender is ambiguous, and consequently, they are slower processed than masculine epicenes, for example. It should be mentioned stereotypical bigenders and neutral bigenders could also be influenced by default gender, but at a less degree than epicenes.

According to Pre-test 2, epicenes and stereotypical bigenders also differ in function of carrying either masculine or feminine genders. Feminine epicenes were conceptually processed faster than masculine epicenes. A reason for that lies in the fact that feminine epicenes are marked in BP, and they may be conceptually more salient than masculine epicenes. On the contrary, masculine stereotypes were

conceptually processed faster than feminine stereotypical bigenders due to the fact world knowledge probabilities might be more easily processed for masculine than feminine stereotypes, as masculine is the default gender in linguistic and in social terms.

Finally, it appears that non-linguistic factors related to factors such as sex of the participants influence the way gender information is retrieved and processed. Male comprehenders tend to see the world with a masculine bias, while feminine comprehenders with a feminine bias, which is in congruence with Kennison & Trofe (2003) and Casado et al. (2017).

Based on the results of Pre-test 1 and Pre-test 2, masculine and feminine epicenes, as well as masculine and feminine stereotypical bigenders were selected for the next experiments. The criteria used was the following: in Pre-test 1, masculine epicenes must have been masculine gender assigned, while feminine epicenes must have been feminine gender assigned; and masculine and feminine stereotypical bigenders must have been gender assigned ambiguously; in Pre-test 2, masculine epicenes feminine epicenes should not be gender biased (which was difficult since, as mentioned before, they showed an unexpected gender bias probably caused by semantic factors), while masculine and feminine stereotypical bigenders should be highly gender biased. It should be noted that some epicenes that were gender assigned ambiguously in Pre-test 1 were not selected. In addition, the reason for excluding gender-biased epicenes is because the intention was to have an ideal grammatical gender, which is typically semantically arbitrary. One can find a list with the Pre-tests results in the appendix. The epicenes selected for the experiments are listed in Table 9:

	Nouns	Gender assigning	Z-scores mean
Feminine Epicenes	*criança* (child)	85% feminine	-0.58
	criatura (creature)	92% feminine	-0.54
	estrela de cinema (movie star)	85% feminine	-0.50
	pessoa (person)	92% feminine	-0.47
	visita (guest)	92% feminine	-0.64
	vítima (victim)	92% feminine	-0.58
Masculine Epicenes	*anjo* (angel)	92% masculine	0.30
	bicho (animal)	92% masculine	0.47
	ente (entity)	64% masculine	0.02
	gênio (genius)	78% masculine	0.69
	indivíduo (individual)	92% masculine	0.19
	ser (being)	92% masculine	0.19

Table 9: Epicenes selected from the Pre-tests

It is worth noticing that all epicenes were animated nouns, but not all of them necessarily refer to humans, as the epicenes *criatura* (creature), *bicho* (animal), and *ser* (being). Epicenes that refer to humans are not numerous in BP, and some of them, especially masculine epicenes, have very low frequency, as for example, *algoz* (executioner), *bóia-fria* (farm worker), *carrasco* (executioner). Besides that, as mentioned before, some of them were not gender assigned as real epicenes, but as bigenders. Thus, it was hard to find to comprise [+human] epicenes, frequency, gender assigning, and gender bias. Although, some epicenes selected do not necessarily refer to humans, in the upcoming experiments, the [+human] interpretation was conveyed in the sentences.

The stereotypical bigenders selected for the experiments are listed in Table 10:

	Nouns	Gender assigning	Z-scores mean
Feminine Stereotypical Bigenders	*assistente social* (social worker)	71% ambiguous	-0.64
	esteticista (beautician)	71% ambiguous	-0.64
	ginecologista (gynecologist)	85% ambiguous	-0.69
	modelo (model)	78% ambiguous	-0.69
	nutricionista (nutricionist)	78% ambiguous	-0.80

	recepcionista (receptionist)	78% ambiguous	-0.75
Masculine Stereotypical Bigenders	*detetive* (detective)	71% ambiguous	0.41
	eletricista (electrician)	57% ambiguous	0.30
	guarda (guard)	50% ambiguous	0.47
	taxista (taxi driver)	71% ambiguous	0.41
	surfista (surfist)	71% ambiguous	0.20
	piloto de corrida (race car driver)	57% ambiguous	0.91

Table 10: Stereotypical bigenders selected from the Pre-tests

One can see that some masculine stereotypical bigenders such as *eletricista* (electrician), *guarda* (guard), and *piloto de corrida* (race car driver) were not highly gender assigned as ambiguous. This happened because some masculine stereotypical bigenders were so masculine biased that they were syntactic gender assigned as masculine.

4.4 Eye-tracking experiments

The main purpose of the eye-tracking experiments was to investigate the role of gender morphological cues in how pronouns retrieve their antecedents in a language with overt morphology. Lago (2014) showed speakers of Spanish are used to rely on morphological cues when processing subject-verb dependencies; therefore, it is hypothesized speakers of Portuguese would also be used to rely on morphological cues when processing coreference dependencies. Therefore, the first hypothesis of this dissertation is gender cues are of paramount importance in antecedent retrieval in memory in languages such as Brazilian Portuguese.

Since the main purpose of this dissertation was to investigate the role of gender morphological cues in antecedent retrieval, it was necessary to dissociate them from structural constraints. Thus, a way for doing that was analyzing the role of

structurally unacceptable antecedents that agree with the pronouns in gender, despite the fact they violate the structural constraints of Principle B.

Few previous studies as Badecker & Straub (2002), Kennison (2003), Cunnings & Felser (2013), Parker (2014) and Patil at al. (2016) found out evidences in favor of a initial fallibility of binding structural constraints, that is, when structurally unacceptable antecedent candidates that agree with the anaphoric expression influence early phases of binding processing, even when there is a structurally acceptable antecedent available in the sentence. This way, it is expected the same would happen in Brazilian Portuguese, that is, structurally unacceptable antecedents that agree with the pronouns in gender would influence antecedent retrieval, but not only at initial processing phases, but also at late processing phases. Since Portuguese is a language with overt and redundant morphology, it is hypothesized that gender cues would play a role in pronominal antecedent retrieval from the beginning to the end of coreference processing.

Furthermore, this dissertation also controlled for the types of gender conveyed by the structurally unacceptable antecedents with the purpose of verifying which types of gender weighed more in memory. This leads to the second hypothesis, which argues memory is able to distinguish the differences that exist between different types of gender, so that different types of gender are encoded/retrieved in memory with different weights (van Dyke & McElree, 2011; Dillon et al., 2013). And because different types of gender would have different types of weights in memory, some of them would be more prominent in memory than others. The concept of prominence in memory is used in this dissertation to refer to grammatical cues that weigh more in memory; it not related to the idea of prominence conveyed by the Discursive Prominence Theory (Gordon & Hendrick, 1998).

Vigliocco & Franck (1999) showed nouns with semantic gender such as *arquiteta* (female arquitect) and *arquiteto* (male architect) are processed more easily than nouns with grammatical gender, which are also called epicenes, such as *vítima* (victim) and *indivíduo* (individual). A possibly reason for that lies in the fact semantic gender is redundant, that is, it is both syntactically and semantically gender specified, while grammatical gender is only syntactically specified. Following this rationale, semantic gender seems to be more prominent than grammatical gender.

On the other hand, definitional gender of whole-forms such as *mulher* (woman) and *homem* (man) would be more prominent in memory than stereotypical gender as in *recepcionista* (receptionist) and *surfista* (surfist). In the latter case, gender information would be probabilistically inferred based on world-knowledge (Oakhill et al. 2005; Carreiras et al., 1996; Kneiner et al., 2008; Canal et al., 2015), which seems more psychologically demanding than retrieving gender from whole-forms with definitional gender as showed by Oakhill et al. (2008) and Canal et al. (2015).

Moreover, when comparing two types of semantic gender, as for example, definitional gender and compositional/derivational gender as in nouns such as *arquiteto* (male architect) and *arquiteta* (female architect), one can expect definitional gender would more prominent than compositional/derivational gender. The logic behind this idea is the dual-route mechanism processing (Ullman et al., 1997; Pinker, 1991; Allen et al., 2003). Gender information of nouns with lexically determined genders would be fully accessed together with the word retrieval in the lexicon, while gender information of nouns with compositional/derivational semantic gender would require a more psychological demanding processing. According to Affix Stripping Hypothesis (Taft & Forster, 1975), the gender suffix of compositional/derivational

nouns would be stripped off so that the noun base form would be accessed and eventually joined with the gender suffix in order to fully process the word.

In other words, it is predicted definitional gender would be more prominent than both stereotypical gender and compositional/derivational gender, which is expected to be more prominent than grammatical gender.

In order to test the differences among the different type of gender, four eye-tracking experiments were conducted in this dissertation. The first two (Experiments 1a and 1b) focused on testing the differences between grammatical gender and compositional/derivational semantic gender in pronominal antecedent retrieval, while the last two (Experiments 2a and 2b) focused on comparing definitional semantic gender and stereotypical gender.

4.4.1 Eye-tracking

Before reporting the eye-tracking experiments, this section will briefly present some main concepts of the eye-tracking methodology.

In eye-tracking experiments, participants have their eye movements recorded while they read text on a computer screen. Using appropriate software, the researcher can measure the duration of eye fixations (among other measures). This technique is one of the most efficient means linguists have to study language processing. Moreover, it has advantages over the self-paced reading technique because the text can be presented more naturally to the readers (i.e, without segmentation and button pressing).

According to Just & Carpenter (1980), the duration of eye fixations during sentence processing depends on information complexity, that is, the more complex

information processing is, the longer the fixation duration in the area where that information is located. These authors make two assumptions: the first is called the *Immediacy Assumption*, which claims that language processing is immediate, that is, a word is processed at the first time it is encountered; the second is called the *Eye-Mind Assumption*, which means that the eye remains fixated on a word as long as the word is being processed. The first assumption is still considered true; however, the second assumption is no longer thought to be true, since a word can still be processed when the eyes are fixated on the next word, which is called the *spillover effect*.

The eye-tracking measures that will be used in the present study are: (a) First Fixation, duration of the first fixation in a word or region of interest; (b) First Pass or Gaze Duration, sum of the durations of all fixations on a word or region before leaving it to the right or to the left; (c) Regression Path or Go-Past Time, corresponds to the duration of all fixations from first fixation on a region to first moving to the right – this included regressions back to earlier parts of the sentence before moving on; and (d) Second Pass, sum of all re-fixations in a region. These measures are considered to be the standard measures in the eye-tracking literature.

On the one hand, First Fixation and First Pass are considered to be early eye-tracking measures, which correspond to the very beginning of processing. With respect to First Fixation, it is considered to depict the time it takes the readers to lexically retrieve a word. On the other hand, Second Pass is considered to be a late eye-tracking measure. With respect to Regression Path, it can reflect either difficulties in integrating information in the context, which can be considered an early effect; or a cost of overcoming a difficulty, which can be considered a late effect in processing.

Since the first hypothesis of this dissertation claims gender morphological cues plays a great role in pronominal antecedent retrieval in Brazilian Portuguese, it is

expected gender cues would influence antecedent retrieval from the beginning to the end of coreference processing. In other words, it is predicted to find pervasive gender cues effects, that is, gender cues effects would be found in each of the eye-tracking measures mentioned above.

4.4.2 Experiments 1a and 1b

Experiments 1a and 1b aimed at investigating the differences between grammatical gender and compositional/derivational semantic gender during antecedent retrieval in Brazilian Portuguese. By comparing these two types of gender, it would be possible to verify whether compositional/derivational semantic gender would weigh more (see section 2.6.5) in memory than grammatical gender. This prediction is based on the finding of Vigliocco & Franck (1999), who showed semantic gender would be more easily processed than grammatical gender since the former is redundantly gender specified. In other words, they claimed semantic gender is gender specified both syntactically and semantically, while grammatical gender is only gender specified syntactically.

Since compositional/derivational semantic gender was predicted to weigh more in memory than grammatical gender, it was expected structurally unacceptable candidates carrying compositional/derivational semantic gender would be responsible for slower coreference processing. A reason for that lies in the fact structurally unacceptable candidates with compositional/derivational gender would be more preferable candidates than the ones with grammatical gender as the former carry semantic gender. A greater preference would mean larger interference effects, that is, more competition with the structurally acceptable antecedents and, consequently, slower coreference processing.

Moreover, since masculine is the default gender (Corbett, 1991; Casado et al., 2017), it is expected for masculine structurally unacceptable candidates to be responsible for slower antecedent retrievals. They would cause greater interference effects than feminine since they are more preferable to be retrieved. In other words, masculine gender would weigh more in memory than feminine gender.

Both Experiments 1a and 1b tested grammatical and compositional/derivational semantic genders; however, Experiment 1a tested those types of gender in the feminine and Experiments 1b in the masculine as illustrated in the table below.

Eye-tracking experiment	Type of gender	Structurally unacceptable antecedent candidates examples
Experiment 1a	Feminine compositional/derivational semantic gender	*psicóloga* (female therapist)
	Feminine grammatical gender	*pessoa* (person)
Experiment 1b	Masculine compositional/derivational semantic gender	engenheiro (male engineer)
	Feminine grammatical gender	gênio (genius)

Table 11: Experiments 1a and 1b

4.4.2.1 Participants

32 native speakers of Brazilian Portuguese (26 female and 10 male, average of age of 22 years) participated in Experiment 1a; and 36 (22 female and 14 male, average of age of 22 years) participated in Experiment 1b. All participants were randomly invited to participate in this experiment as volunteers. They are undergraduate students at the

Federal University of Rio de Janeiro (Rio de Janeiro, Brazil) and have normal or corrected-to-normal vision. All participants were naive in relation to the object of study of the experiment and signed a consent form giving permissions to the experimenter to publish the results. They received 3 hours of Cultural-Scientific Activities (*Atividades-Científico-Culturais Discentes*, AACC) as compensation for their work.

4.4.2.2 Materials and design

The experimental materials of each experiment consisted of 48 sentences distributed in 4 conditions. The experimental trials were arranged into 4 lists using a Latin Square. Each list was pseudo-randomized and contained 12 experimental items and 24 fillers. Each and every trial was accompanied by a comprehension question. The experimental trials were composed by embedded third-person-singular pronouns (*ele/ela*) with pronominal antecedents (masculine/feminine common nouns) followed by distractors, which are close antecedent candidates that cannot be considered as structurally acceptable antecedents due to Principle B structural constraints. Thus the structurally acceptable antecedents are the preferable antecedent candidates not only due to structural constraints, but also due to discursive factors. They are the subjects of the main clause, which means they are highly accessible, and they can be considered as the discursive topic. However, the structurally unacceptable antecedent candidate is the subject of the embedded clause, and even though they are not as prominent as the structurally acceptable antecedents, they are accessible in discourse due to its recentness.

The experimental trials were composed by 4 regions of interest. The critical region was the *pronoun region*, which contained the pronouns *ele* (him) or *ela* (her),

formed by 3 characters. Before the *pronoun region*, there were a relative pronoun *que* (who), which introduces the relative clause, followed by a transitive verb (approximately 5-6 characters). After the *pronoun region*, there were 3 spillover regions – the *pronoun +1 region*, the *pronoun +2 region*, and the *pronoun +3 region*. The *pronoun +1 region* contained an adverb of manner (approximately 9-11 characters), the pronoun +2 region contained a prepositional or adverbial phrase (approximately 5-9 characters), and pronoun +3 region contained a noun or a propositional phrase (approximately 8-12 characters). The word size control is important in reading experiments as an average of the reading times at the same region of interest across the materials will be calculated. The regions of interest need to be as similar as possible in word size. This way, the differences in reading times that would eventually appear among the same region of interest would be provoked by a condition effect, and not by word size.

Since the critical region is too small (only 3 characters), it expected that any processing difficulties readers could have at the pronoun region would spread to subsequent regions, which are the spillover regions (see section 4.4.1). Therefore, the 3 spillover regions used in the eye-tracking experiments of this dissertation could capture late reading effects generated at the pronoun region.

The independent variables of the experiment are: a) *antecedent matching the gender of the pronoun*, which is a factor that is directly related to Principle B; and b) *distractor matching the gender of the pronoun*, which is factor that relies purely on agreement cues. This way, the experimental design was 2x2, with four main conditions. We also controlled for the *distractor type of gender*; therefore, half of the experimental trials contained distractors with semantic gender and the other half contained distractors with grammatical gender. It should be noticed that all distractors

of Experiment 1a were feminine and all distractors of Experiment 1b were masculine. A sample of the materials of Experiment 1a can be found in Tables 11 and 12 and a sample of the materials of Experiment 1b can be found in Tables 13 and 14. Brackets delimit the regions of interest.

(82) Distractor with feminine semantic gender:

	Antecedent mismatch	**Antecedent match**
Distractor match	O bailarino admira a psicóloga que ajudou [**ela**] [gentilmente] [depois] [de uma das] fases mais difíceis na vida. (*The dancer$_{[masc]}$ admires the therapist$_{[fem]}$ who gently helped **her** after one of the most difficult phases in life.)*	A bailarina admira a psicóloga que ajudou [**ela**] [gentilmente] [depois] [de uma das] fases mais difíceis na vida. (*The dancer$_{[fem]}$ admires the therapist$_{[fem]}$ who gently helped **her** after one of the most difficult phases in life.)*
Distractor mismatch	A bailarina admira a psicóloga que ajudou [**ele**] [gentilmente] [depois] [de uma das] fases mais difíceis na vida. (*The dancer$_{[fem]}$ admires the therapist$_{[fem]}$ who gently helped **him** after one of the most difficult phases in life.)*	O bailarino admira a psicóloga que ajudou [**ele**] [gentilmente] [depois] [de uma das] fases mais difíceis na vida. (*The dancer$_{[masc]}$ admires the therapist$_{[fem]}$ who gently helped **him** after one of the most difficult phases in life.)*

Table 12: Sample of the materials for distractors with feminine semantic gender used in Experiment 1a by regions of interest

(83) Distractor with feminine grammatical gender:

	Antecedent mismatch	Antecedent match
Distractor mismatch	O veterinário reconheceu a pessoa que feriu [**ela**] [fortemente] [por trás] [da cabeça] no momento do assalto. (*The vet$_{[masc]}$ recognized the person$_{[fem]}$ who heavily hurt **her** behind the head at the moment of the robbery.)*	A veterinária reconheceu a pessoa que feriu [**ele**] [fortemente] [por trás] [da cabeça] no momento do assalto. (*The vet$_{[fem]}$ recognized the person$_{[fem]}$ who heavily hurt **her** behind the head at the moment of the robbery.)*
Distractor match	A veterinária reconheceu a pessoa que feriu [**ele**] [fortemente] [por trás] [da cabeça] no momento do assalto. (*The vet$_{[fem]}$ recognized the person$_{[fem]}$ who heavily hurt **him** behind the head at the moment of the robbery.)*	O veterinário reconheceu a pessoa que feriu [**ele**] [fortemente] [por trás] [da cabeça] no momento do assalto. (*The vet$_{[masc]}$ recognized the person$_{[fem]}$ who heavily hurt **him** behind the head at the moment of the robbery.)*

Table 13: Sample of the materials for distractors with feminine grammatical gender used in Experiment 1a by regions of interest

(84) Distractor with masculine semantic gender:

	Antecedent mismatch	Antecedent match
Distractor match	A arquiteta agradeceu o engenheiro que indicou [**ele**] [justamente] [para um] [dos cargos] mais cobiçados do país. (*The arquitect$_{[fem]}$ thanked the engineer$_{[masc]}$ who fairly recommended **him** for one of the most desirable jobs in the country.)*	O arquiteto agradeceu o engenheiro que indicou [**ele**] [justamente] [para um] [dos cargos] mais cobiçados do país. (*The arquitect$_{[mas]}$ thanked the engineer$_{[masc]}$ who fairly recommended **him** for one of the most desirable jobs in the country.)*
Distractor mismatch	O arquiteto agradeceu o engenheiro que indicou [**ela**] [justamente] [para um] [dos cargos] mais cobiçados do país. (*The arquitect$_{[mas]}$ thanked the engineer$_{[masc]}$ who fairly recommended **her** for one of the most desirable jobs in the country.)*	A arquiteta agradeceu o engenheiro que indicou [**ela**] [justamente] [para um] [dos cargos] mais cobiçados do país. (*The arquitect$_{[fem]}$ thanked the engineer$_{[masc]}$ who fairly recommended **her** for one of the most desirable jobs in the country.)*

Table 14: Sample of the materials for distractors with masculine semantic gender used in Experiment 1b by regions of interest

(85) Distractor with masculine grammatical gender:

	Antecedent mismatch	Antecedent match
Distractor match	A milionária recompensou o gênio que alertou [**ele**] [severamente] [a respeito] [dos efeitos] da crise na economia. (*The millionaire$_{[fem]}$ rewarded the genius$_{[masc]}$ who severely alerted **him** about the effects of the crisis in economy.)*	O milionário recompensou o gênio que alertou [**ele**] [severamente] [a respeito] [dos efeitos] da crise na economia. (*The millionaire$_{[masc]}$ rewarded the genius$_{[masc]}$ who severely alerted **him** about the effects of the crisis in economy.)*
Distractor	O milionário recompensou o gênio	A milionária recompensou o gênio

mismatch	que alertou [**ela**] [severamente] [a respeito] [dos efeitos] da crise na economia. (*The millionaire[masc] rewarded the genius[masc] who severely alerted **her** about the effects of the crisis in economy.)*	que alertou [**ela**] [severamente] [a respeito] [dos efeitos] da crise na economia. (*The millionaire[fem] rewarded the genius[masc] who severely alerted **her** about the effects of the crisis in economy.)*

Table 15: Sample of the materials for distractors with masculine grammatical gender used in Experiment 1b by regions of interest

The on-line dependent variables for both Experiments 1a and 1b are the following reading measures at the pronoun and at the spillover regions: (1) First Fixation Duration; (2) First Pass; (3) Regression; (4) Regression Path; (5) Right Bound; (6) Second Pass; and (7) Total Time[18].

4.4.2.3 Procedure

The experiment was conducted at the laboratory of experimental research (LAPEX) at the Federal University of Rio de Janeiro (UFRJ) in Rio de Janeiro, Brazil. The eye-tracker used in this experiment was *Eye Link 1000* and the experiment was programmed and conducted on *Eye Track 7.10m* [19]software. All trials were typed in font *Monaco* size 12. The participants were instructed to seat comfortable and were

[18] See section 4.1.1 to have a more detailed explanation on eye-tracking measures.

[19] The primary developers of *Eye Track* were David Stracuzzi and Jeff Kinsey and it is conceptually based on software written by Saarbruken and provided to UMASS by Christoph Scheepers. *Eye Track* can be downloaded for free on https://blogs.umass.edu/eyelab/software/.

given written and oral task instructions. The instructions screen is illustrated in Figure 17[20].

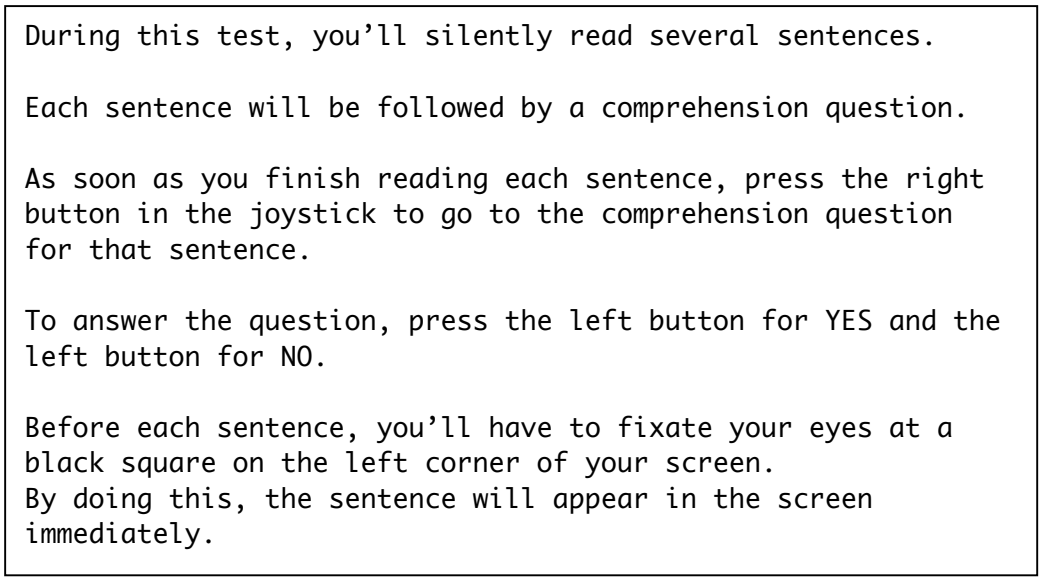

Figure 17: Instructions screen of Experiments 1a and 1b

After receiving the instructions, the calibration process would start followed by a short practice with 6 filler sentences so that the experimenter would check whether the participants understood the task and were performing it at a natural speed. Each participant performed one of the 4 lists of the experiment, which were pseudo-randomized by *Eye Track* software. The experiment duration was of 20 minutes approximately.

4.4.2.4 Analysis

[20] The participants received the instructions in Portuguese, but we translated them to English for the purpose of this dissertation.

The eye-tracking data was analyzed using the following tools: *Visual EDF to ASC*, to convert the .EDF files that *Eye Link 1000* generates; *Robodoc.py*[21], to clean eye blinks and long saccades (longer than 80ms); *Question_acc.py* [22] to compute the comprehension questions accuracy; *EyeDry*[23] to compute the reading measures; and R for the statistics.

Some experimental trials had to be excluded due to eye blinks and long saccades at the regions of interest (15% in Experiment 1a, and 21% in Experiment 1b). Moreover, 2 participants were excluded from analysis due to very slow reading as they trespassed the time limit in all trials, including fillers. Thus 21% of the experimental data in had to be excluded in Experiment 1a.

Experiment 1a and 1b were put together in as a *between-subjects* analysis. Linear mixed effects models (*lmes*) with random intercepts[24] were created for each region of interest: the critical region (pronoun) and the spillover regions (pronoun+1, pronoun+2, and pronoun+3). The fixed effects of the *lmes* were: a) *antecedent matching the gender of the pronoun* (match/ mismatch), b) *distractor matching the gender of the pronoun* (match/mismatch), the c) *distractor type of gender* (semantic or grammatical); and the d) *distractor gender* (masculine and feminine). On the other hand, the random effects were the *participants* and the *items.*

In order to obtain the F-ratios and the p-values, ANOVAs were applied to the *lmes* to figure out whether they were significantly relevant. *Tukey Honest Significant*

[21] *Rododoc.py* is a python script created by Adrian Staub and Chuck Clifton, and the 2016 version was revised by Jesse Harris. It can also be downloaded on https://blogs.umass.edu/eyelab/software/.

[22] *Question_acc.py* is a python script that comes with Robodoc.py utils to check questions accuracy and their reaction times.

[23] *EyeDry* was created by Chuck Clifton and can be downloaded on https://blogs.umass.edu/eyelab/software/.

[24] *Lmes* with random slopes did not converge.

Differences (HSD) tests were also conducted as pairwise comparison post-hoc tests with the purpose of checking which conditions were statistically different.

Since the ANOVAs results found statistically significant interactions for two factors or three fixed factors. Interaction plots were created using the *plot ( )* function of *effects*[25] package in R with the objective of clarifying the relationship between the factors involved in those interactions.

4.4.2.5 Results

It should be mentioned that the participants answered the comprehension questions with an average of accuracy of 88% in Experiment 1a and 93% in Experiment 1b, which means that the participants were paying attention to the task and reading the sentences properly.

Results for each of the eye-tracking measures will be reported and discussed below according to the regions of interest investigated.

First Fixation

For space reasons, only the statistically significant results will be reported in this section.

Figure 18 corresponds to the First Fixation at all regions of interest in Experiments 1a and 1b.

[25] John Fox, Jangman Hong (2009). Effect displays in R for Multinomonial and Proportional-odds Logit Models: Extentions to the effects Package. Journal of Statistical Software, 32 (1), 1-24. URL http://www.jstatsoft.org/v32/i01/.

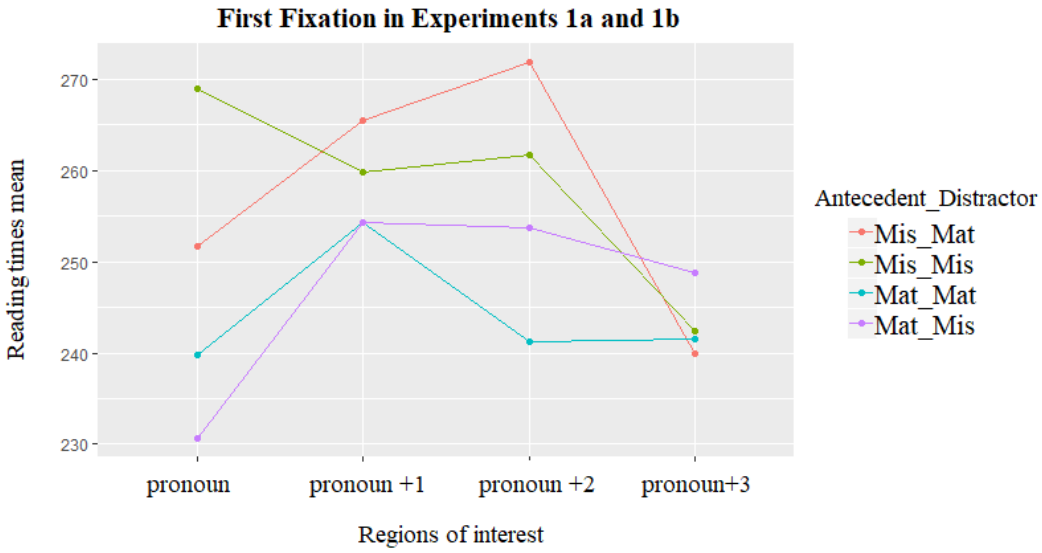

Figure 18: Line chart for First Fixation by regions of interest and conditions in Experiments 1a and 1b

By running Tukey HSD tests, it was observed a statistically significant difference between the conditions Mis_Mis and Mat_Mis at the pronoun region and a marginal statistically significant difference between the conditions Mis_Mat and Mat_Mat at the pronoun+2 region. In other words, it does not matter whether the distractors match the pronouns or not, reading times at the pronoun region are faster when the antecedent match the pronouns in gender. Antecedent retrieval is 38ms (p=0.038) and 30ms (p=0.06) faster with mismatching distractors and matching distractors conditions respectively.

ANOVA tests only found statistically significant results for the pronoun region and pronoun +2 regions. This is the reason why the results of only these two regions will be reported and discussed here.

Pronoun region

The reading times and standard deviations at the pronoun region in milliseconds are reported in Table 15. Tukey HSD tests did not show any statistically significant

pairwise comparisons across the different conditions in the experiment. However, it shows reading times at the pronoun region are 19ms faster for sentences with matching antecedents than for sentences with mismatching antecedents with a trend towards statistical significance (p=0.057).

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	208 (83)	289 (93)	248 (86)	251 (172)
Mis_Mis	270 (89)	230 (82)	279 (103)	295 (164)
Mat_Mat	221 (70)	242 (97)	268 (108)	237 (93)
Mat_Mis	253 (99)	236 (75)	213 (49)	217 (45)

Table 16: First Fixation reading times and standard deviations in parenthesis at the pronoun regions in Experiments 1a and 1b

ANOVAs of the *lmes* for First Fixation at the pronoun region revealed a statistically significant interaction between *antecedent*, *distractor* and *type of distractor gender*: $F(1, 9026) = 3.94$, $p = 0.047$, and a statistically significant interaction between *distractor*, *type of distractor gender,* and *distractor gender*: $F(1, 9026) = 6.87$, $p = 0.009$. Interaction plots were created in order to explain these interactions.

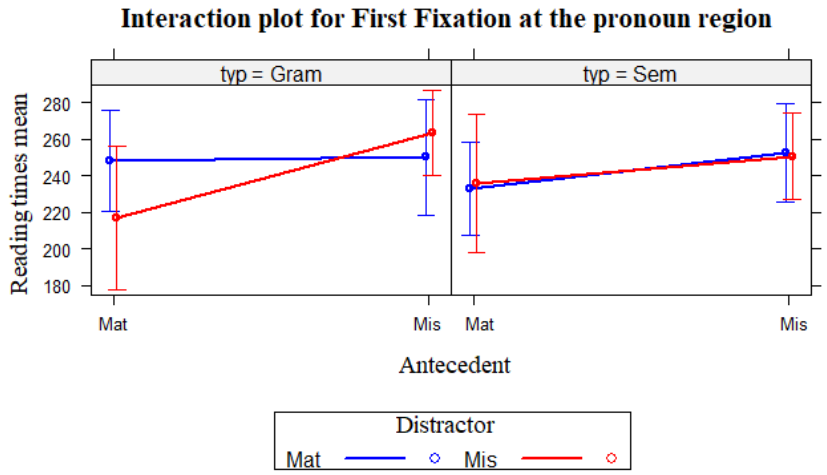

Figure 19: Interaction plot for First Fixation at the pronoun region by antecedent, distractor, and type of gender in Experiment 1a and 1b

Tukey HSD tests did not find any statistically significant differences in Figure 19.

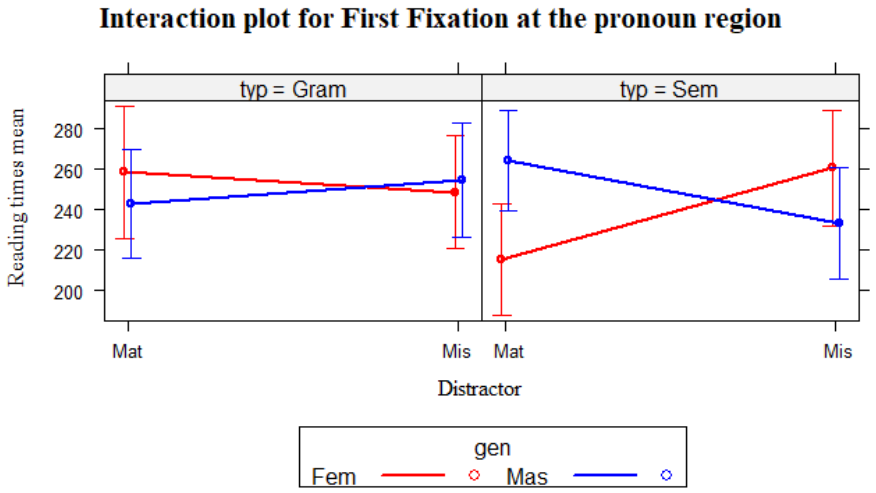

Figure 20: Interaction plot for First Fixation at the pronoun region by distractor, type of gender, and gender in Experiments 1a and 1b

Although it was a trend towards significance, Tukey HSD tests indicated reading times at the pronoun region is 47ms slower with a matching antecedent

followed by a distractor with masculine semantic gender than with a distractor with feminine semantic gender (p=0.189).

Pronoun +2 region

The reading times and standard deviations at the pronoun +2 region in milliseconds are reported in Table 16. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment. However, they show reading times at the pronoun+2 region for sentences with matching antecedents are 19ms faster than for mismatching antecedents (p=0.032).

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	304 (169)	280 (132)	246 (108)	252 (90)
Mis_Mis	252 (91)	283 (111)	261 (95)	247 (121)
Mat_Mat	234 (59)	256 (93)	265 (108)	222 (73)
Mat_Mis	245 (92)	255 (99)	280 (102)	248 (84)

Table 17: First Fixation reading times and standard deviations in parenthesis at the pronoun +2 region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for First Fixation at the pronoun+2 region indicated a statistically significant main effect of *antecedent*: F(1, 9255)=4.14, p=0.042. In addition, it was found a series of statistically significant interactions between *antecedent* and *distractor type of gender*: F(1, 9255)=4.02, p=0.045; between *distractor* and *distractor type of gender*: F(1,9255)=4.54, p=0.033; and between *antecedent*, *distractor type of gender*, and *distractor gender*: F(1, 9255)=3.95, p=0.047. Interaction plots were made for clarifying the interactions.

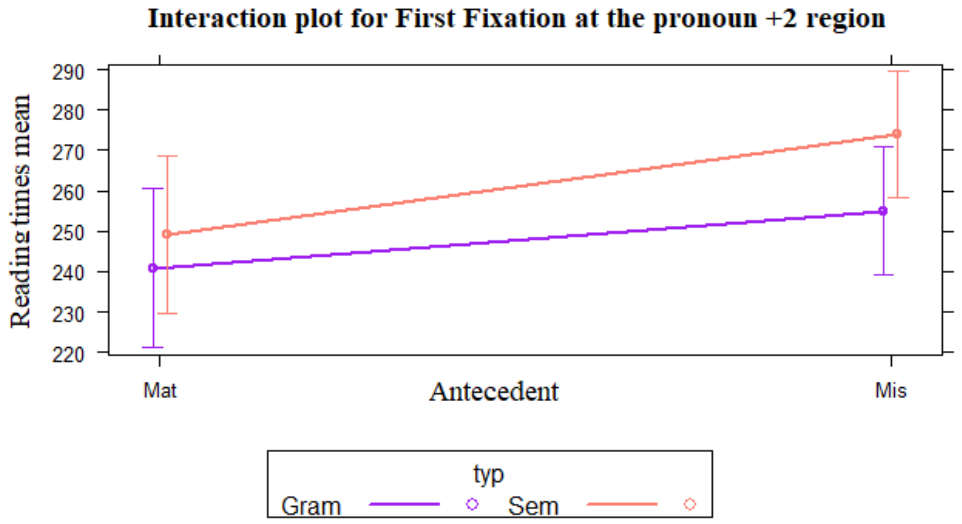

Figure 21: Interaction plot for First Fixation at the pronoun +2 region by antecedent and type of gender in Experiments 1a and 1b

Tukey HSD tests showed reading times at the pronoun +2 region is 32ms longer for mismatching antecedents followed by distractors with semantic gender than for matching antecedents followed by distractors with grammatical gender (p=0.050).

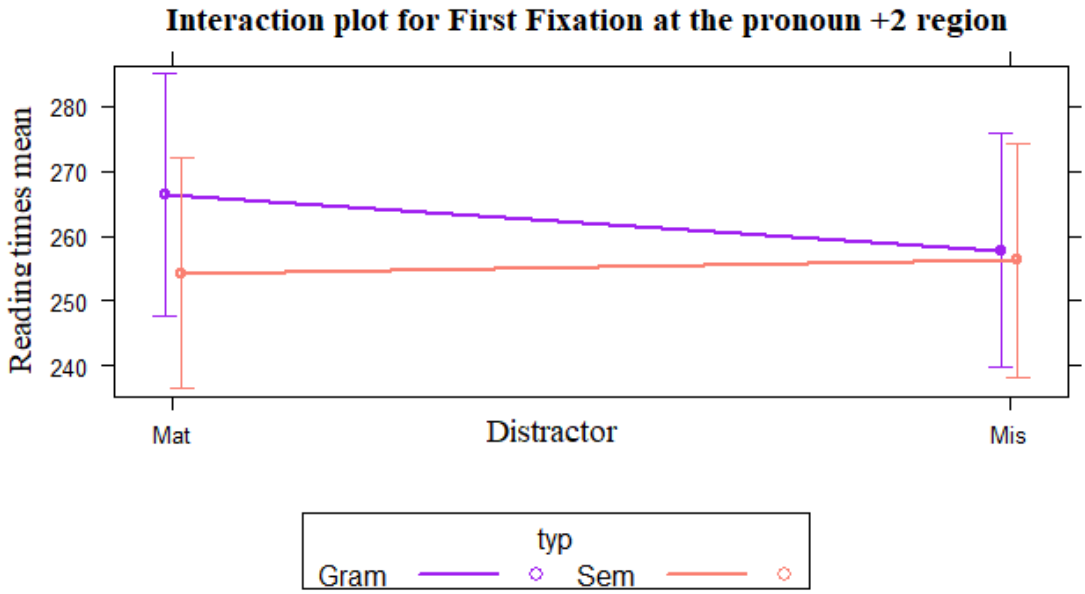

Figure 22: Interaction plot for First Fixation at the pronoun +2 region by distractor and type of gender in Experiments 1a and 1b

Although it was a trend towards significance, Tukey HSD tests showed reading times at the pronoun +2 region is 26ms longer for matching distractors with semantic gender than for matching distractors with grammatical gender (p=0.154).

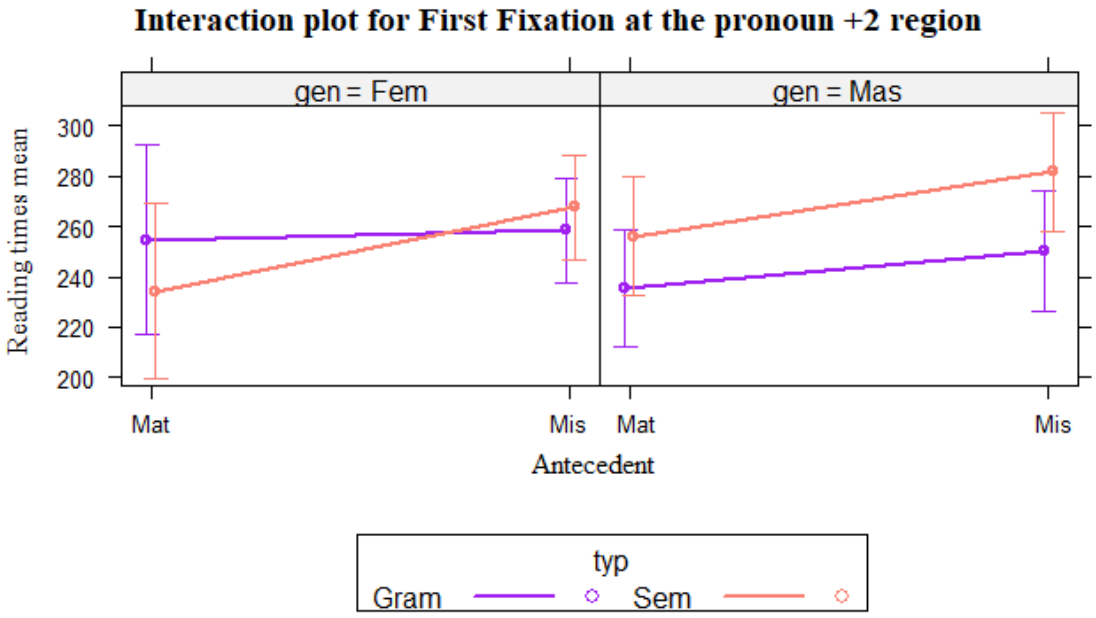

Figure 23: Interaction plot for First Fixation at the pronoun +2 region by antecedent, type of gender, and gender in Experiments 1a and 1b

Tukey HSD tests showed reading times at the pronoun +2 region is 47ms longer when mismatching antecedents are followed by distractors with masculine semantic gender than when matching antecedents are followed by distractors with masculine grammatical gender in a trend towards significance (p=0.086).

Discussion

First Fixation corresponds to the duration of the very first fixation at a region of interest. It is the earliest eye-tracking measure and it frequently related to lexical retrieval. Thus finding positive results for this reading measure is piece of evidence that antecedent retrieval in memory starts as soon as the pronoun in encountered. The results of the present work detected that both the structural cues of Principle B and the morphological cues play a role in First Fixation.

As already predicted, semantic gender and masculine gender would be more preferably retrieved, that is, would weigh more in memory, when compared to grammatical gender and feminine gender, due to the fact the semantic gender is conceptually motivated and the masculine gender is a default gender in the language.

First Pass

Figure 24 corresponds to the First Pass at all regions in Experiments 1a and 1b. There were no differences for First Pass between the conditions.

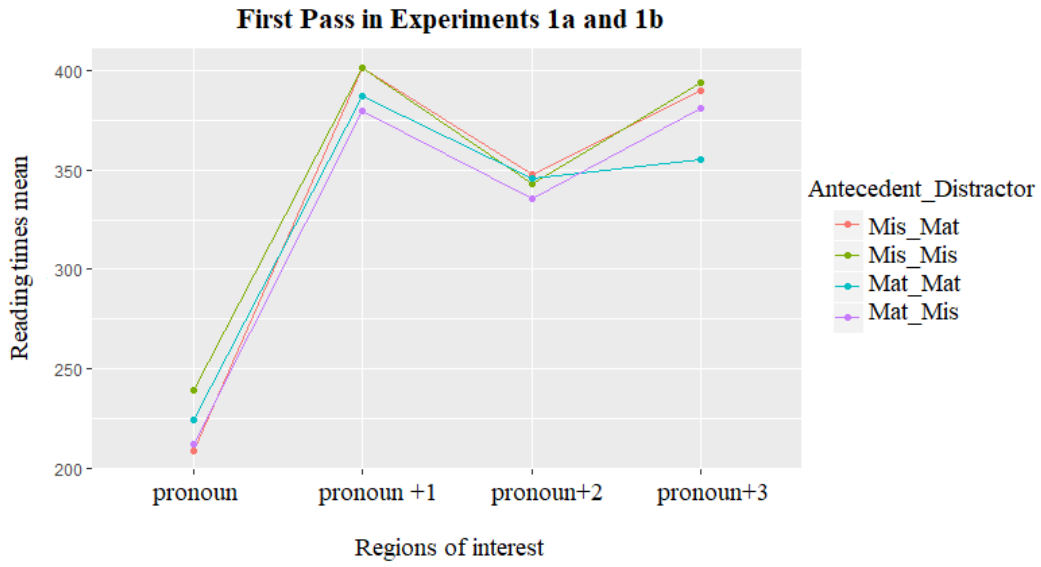

Figure 24: First Pass by regions of interest and conditions in Experiments 1a and 1b

Pronoun region

The reading times and standard deviations at the pronoun region in milliseconds are reported in Table 17. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment.

However, they show masculine distractors speeds up the reading times in 92ms ($p<0.05$) when compared to feminine distractors.

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	156 (146)	300 (99)	140 (143)	284 (204)
Mis_Mis	194 (201)	265 (135)	227 (179)	295 (164)
Mat_Mat	188 (150)	265 (143)	145 (122)	265 (134)
Mat_Mis	209 (167)	272 (131)	181 (164)	239 (74)

Table 18: First Pass reading times and standard deviations in parenthesis at the pronoun region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for First Pass at the pronoun region indicated a statistically significant main effect of *distractor gender*: $F(1, 21085)=31.33$, $p<0.05$.

Pronoun +1 region

The reading times and standard deviations at the pronoun +1 region in milliseconds are reported in Table 18.

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	357 (205)	399 (283)	440 (225)	409 (330)
Mis_Mis	456 (319)	410 (231)	337 (208)	399 (271)
Mat_Mat	327 (160)	378 (213)	381 (187)	394 (202)
Mat_Mis	286 (162)	392 (218)	463 (332)	437 (226)

Table 19: First Pass reading times and standard deviation in parenthesis at the pronoun +1 region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for First Pass at the pronoun+1 region showed a statistically significant interaction between antecedent and distractor type: F(1, 51046)=4.41, p=0.035. The interaction plot in Figure 25 explains this interaction.

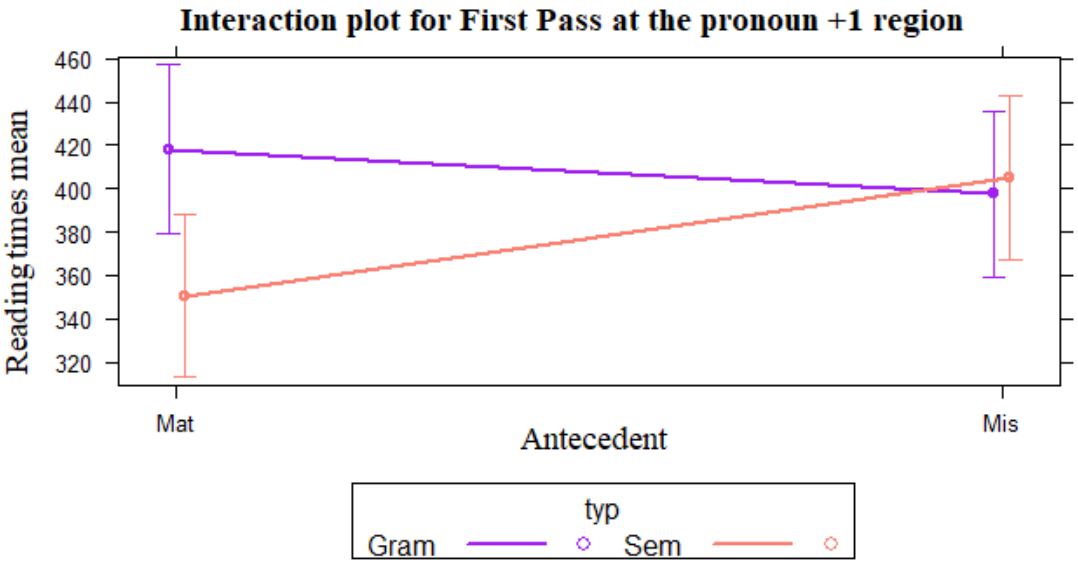

Figure 25: Interaction plot for First Pass at the pronoun +1 region by antecedent and type of gender in Experiments 1a and 1b

Tukey HSD tests revealed reading times at the pronoun +1 region are 67ms longer with matching antecedents followed by distractors with grammatical gender than matching antecedents followed by distractors with semantic gender in a trend towards significance (p=0.069). In addition, reading times at the pronoun +1 region are 54ms faster for matching antecedents followed by distractors with semantic gender than for mismatching antecedents with semantic gender in a trend towards significance (p=0.190).

Pronoun +2 region

The reading times and standard deviations at the pronoun +2 region in milliseconds are reported in Table 19.

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	398 (236)	377 (228)	318 (250)	296 (133)
Mis_Mis	332 (254)	387 (231)	371 (234)	300 (174)
Mat_Mat	301 (294)	423 (213)	376 (188)	283 (170)
Mat_Mis	309 (208)	355 (218)	382 (164)	302 (125)

Table 20: First Pass reading times and standard deviations in parenthesis at the pronoun +2 region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for First Pass at the pronoun+2 region showed a statistically significant interaction between distractor type and distractor gender: $F(1, 35581)=12.46$, $p<0.05$. The interaction plot in Figure 26 aims at clarifying this interaction.

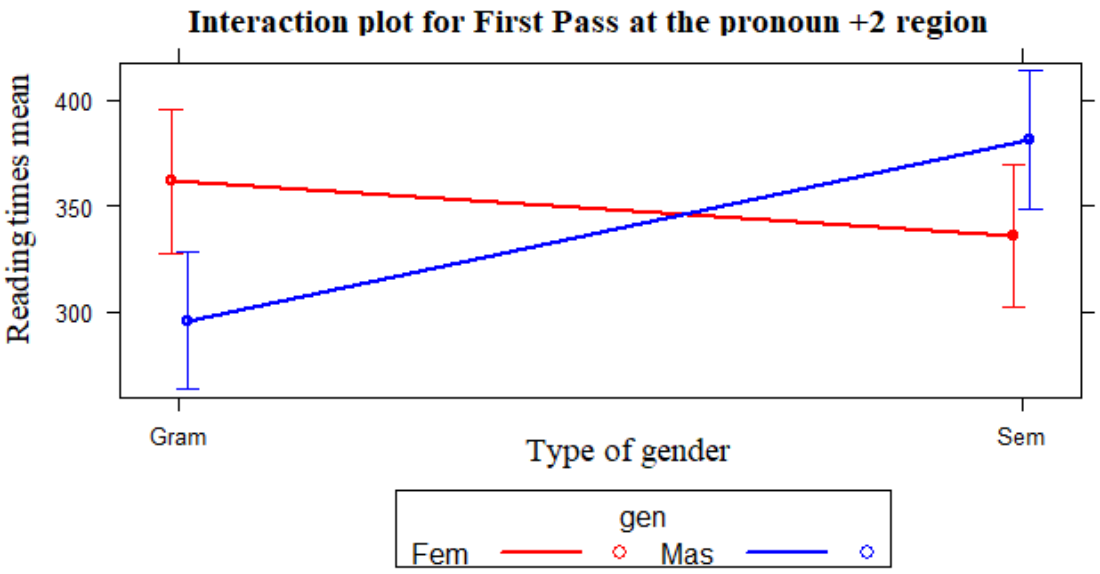

Figure 26: Interaction plot for First Pass at the pronoun +2 region by type of gender and gender in Experiments 1a and 1b

It was found in the Tukey HSD tests that reading times at the pronoun +2 region are 65 ms faster for distractors with masculine grammatical gender than with distractors with feminine grammatical gender (p=0.03). In addition, reading times for masculine distractors with grammatical gender are 24ms faster than masculine distractors with semantic gender (p=0.001).

Pronoun +3 region

The reading times and standard deviations at the pronoun +3 region in milliseconds are reported in Table 20. It was detected in Tukey HSD tests that distractors with semantic gender slow down reading times at the pronoun +3 region in 49ms when compared to distractors with grammatical gender (p=0.010).

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	490 (320)	407 (238)	329 (176)	335 (214)

Mis_Mis	416 (153)	416 (261)	413 (252)	337 (190)
Mat_Mat	349 (170)	371 (219)	296 (159)	327 (164)
Mat_Mis	257 (157)	421 (198)	379 (137)	424 (268)

Table 21: First pass reading times and standard deviations in parenthesis at the pronoun +3 region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for First Pass at the pronoun+3 region showed a statistically significant main effect of *distractor type*: F(1, 41090)=9.09, p=0.002, and statistically significant interactions between *antecedent* and *distractor gender*: F(1, 41090)=9.05, p=0.002, between *antecedent* and *distractor gender*: F(1, 41090)=6.53, p=0.010, and between *antecedent*, *distractor*, *distractor type* and distractor gender: F(1, 41090)=4.28, p=0.038. Interaction plots below illustrate these interactions.

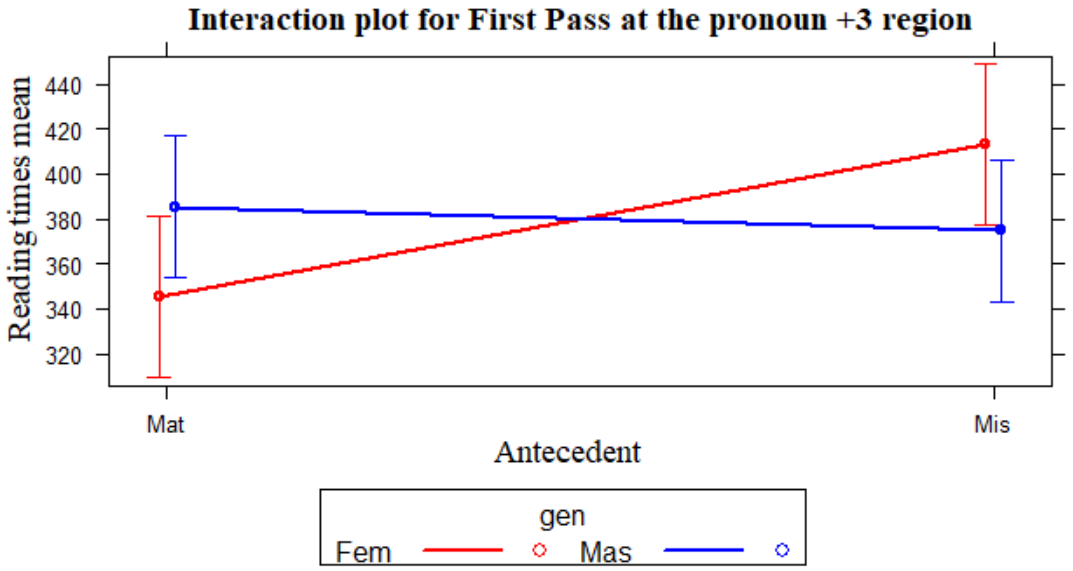

Figure 27: Interaction plot for First Pass at the pronoun +3 region by antecedent and gender in Experiments 1a and 1b

Tukey HSD test shows that reading times at the pronoun +3 region are 68 ms faster for feminine distractors when the antecedents match the pronouns than when they mismatch (p=0.036).

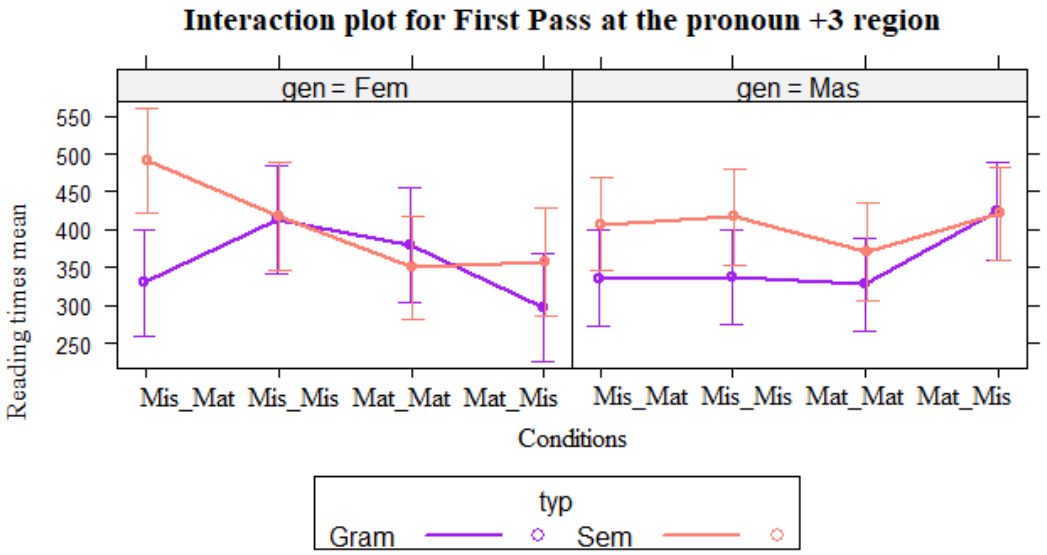

Figure 28: Interaction plot for First Pass at the pronoun +3 region by conditions, type of gender and gender in Experiments 1a and 1b

Tukey HSD tests revealed reading times at the condition Mis_Mat is 161ms longer for feminine semantic gender than for feminine grammatical gender in a trend towards significance (p=0.102).

Discussion

First Pass corresponds to the sum of all the first fixations before the eye exits a region of interest. It is considered to be an early processing measure just like First Fixation Duration. The results for First Pass in Experiments 1a and 1b suggest that at this point, both structural and the morphological cues play a role in antecedent retrieval.

Matching antecedents followed by distractors with grammatical gender had slower antecedent retrievals than followed by distractors with semantic gender. What happens in this case is that distractors with grammatical gender cause similarity-based

interference effects, which is an inhibitory effect predicted by Lewis & Vasishth (2005). Besides that, based on Engelmann et al. (2015), facilitatory effects caused by distractors with semantic gender might mean they are being considered very high prominent in discourse. Thus, the intrusion effects caused by distractors with semantic gender are evidence in favor of the fact they are being retrieved in memory as if they were the antecedents (Dillon et al., 2013). Similar results were found by Cunnings & Felser (2013) and Sturt (2003).

According to Engelmann et al. (2015), similarity-based interference effects can also happen in mismatch-target conditions. When antecedents mismatched the pronouns, distractors with feminine semantic gender slow down pronoun processing when compared to distractors with feminine grammatical gender. As predicted, semantic gender is preferably retrieved in memory than grammatical gender, and since it weighed more than grammatical gender, semantic gender causes more cue confusion as it will compete with the mismatching antecedent.

This way, compared to grammatical gender, semantic gender can either facilitate or inhibit antecedent retrievals. The fact is it weighed more in memory, which can cause either intrusion or cue confusion effects. It all depends on the context involved.

As predicted, masculine distractors are preferably retrieved than feminine gender. Masculine gender weighed more in memory than feminine, because it is the default gender. Thus, distractors with masculine gender have facilitatory effects, that is, intrusion effects in antecedents retrievals. Thus chances are they are being retrieved by memory as antecedents.

Regression Path

Figure 29 corresponds to the Regression Path at all regions in Experiments 1a and 1b.

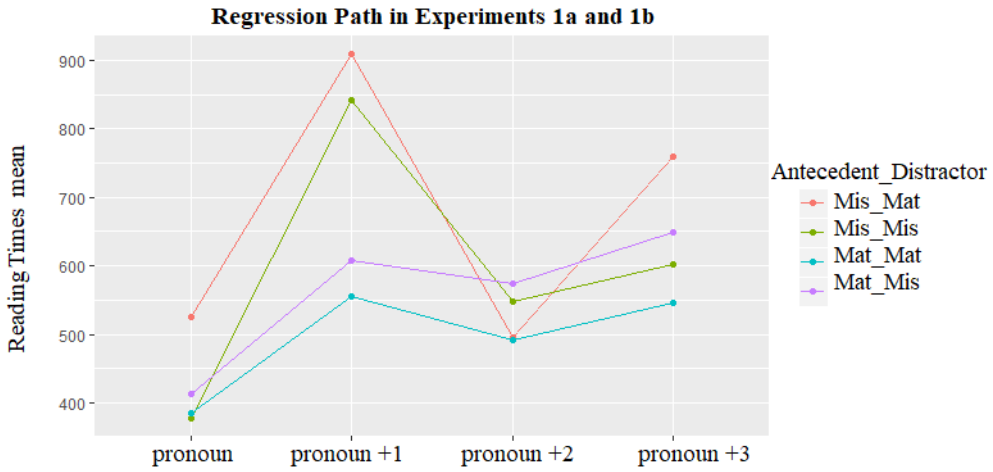

Figure 29: Line chart for Regression Path by regions of interest in Experiments 1a and 1b

It was detected in Tukey HSD tests that at the pronoun region +2, Tukey HSD tests results indicate that condition Mis_Mat is 353ms longer than condition Mat_Mat ($p<0.05$); condition Mis_Mat is 301ms longer than condition Mat_Mis ($p<0.005$); condition Mis_Mis is 286ms longer than condition Mat_Mat ($p=0.002$); and condition Mis_Mis is 234ms longer than condition Mat_Mat ($p=0.016$).

ANOVA did not find statistically significant results for the pronoun region. This is the reason why the results of this region will not be reported and discussed here.

Pronoun +1 region

The reading times and standard deviations at the pronoun +1 region in milliseconds are reported in Table 22. According to the results of Tukey HSD tests, reading times at the pronoun +1 region are 294ms longer when antecedents mismatch the pronouns in gender in comparison to antecedents that match the pronouns ($p<0.05$). In addition, masculine distractors slow down the reading times at the pronoun +1 region in 146ms in comparison to feminine distractors ($p=0.009$).

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	810 (1018)	850 (902)	991 (737)	988 (1014)
Mis_Mis	654 (520)	1015 (966)	593 (373)	1000 (823)
Mat_Mat	468 (334)	528 (491)	631 (409)	610 (666)
Mat_Mis	439 (278)	712 (783)	592 (407)	643(287)

Table 22: Regression Path reading times and standard deviation in parenthesis at the pronoun +1 region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for Regression Path at the pronoun+1 region revealed a statistically significant main effects of *antecedent*: $F(1, 430023)=30.35$, $p<0.05$ and distractor gender: $F(1, 430023)=4.11$, $p=0.046$. In addition, it was found a statistically significant interaction between *distractor* and *distractor gender*: $F(1, 430023)=6.09$, $p=0.013$. The interaction plot below has the purpose of illustrating this interaction.

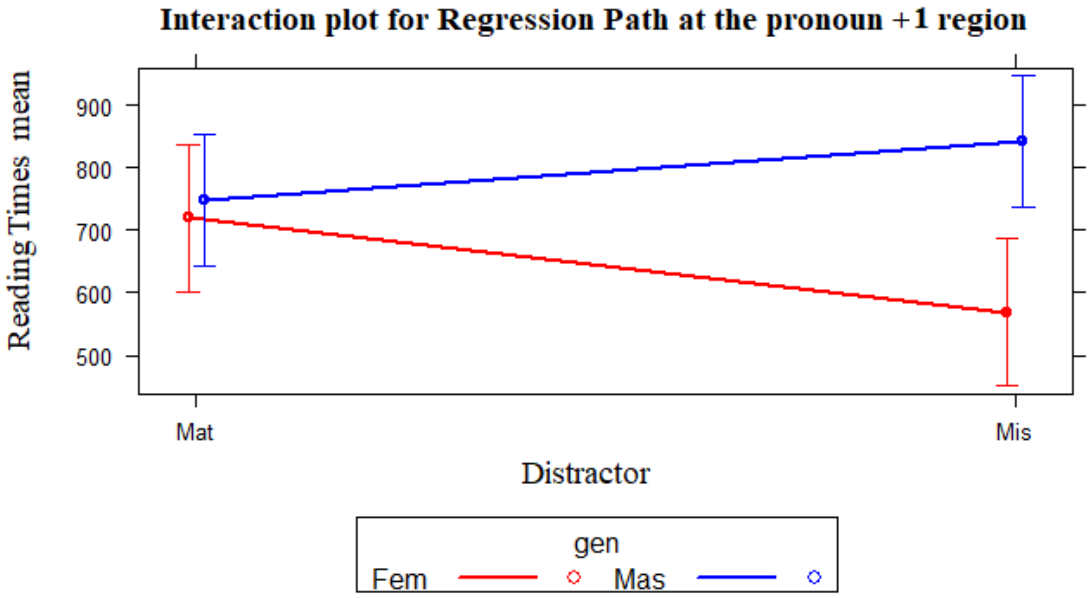

Figure 30: Interaction plot for Regression Path at the pronoun +1 region by distractor and gender in Experiments 1a and 1b

In Figure 30, it is observed feminine mismatching distractors are responsible for 272ms faster reading times than masculine mismatching distractors (p=0.003), according to Tukey HSD tests.

Pronoun +2 region

The reading times and standard deviations at the pronoun +2 region in milliseconds are reported in Table 22.

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	410 (229)	651 (601)	440 (383)	455 (524)
Mis_Mis	595 (496)	536 (355)	409 (276)	652 (958)
Mat_Mat	347 (149)	496 (330)	634 (689)	504 (397)
Mat_Mis	534 (761)	565 (738)	477 (340)	709 (1318)

Table 23: Regression Path reading times and standard deviations in parenthesis at the pronoun +2 region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for Regression Path at the pronoun+2 region revealed a statistically significant interaction between *distractor*, *distractor type*, and *distractor gender*: F(1, 354061)=6.06, p=0.014. The interaction plot in Figure 31 portraits this interaction.

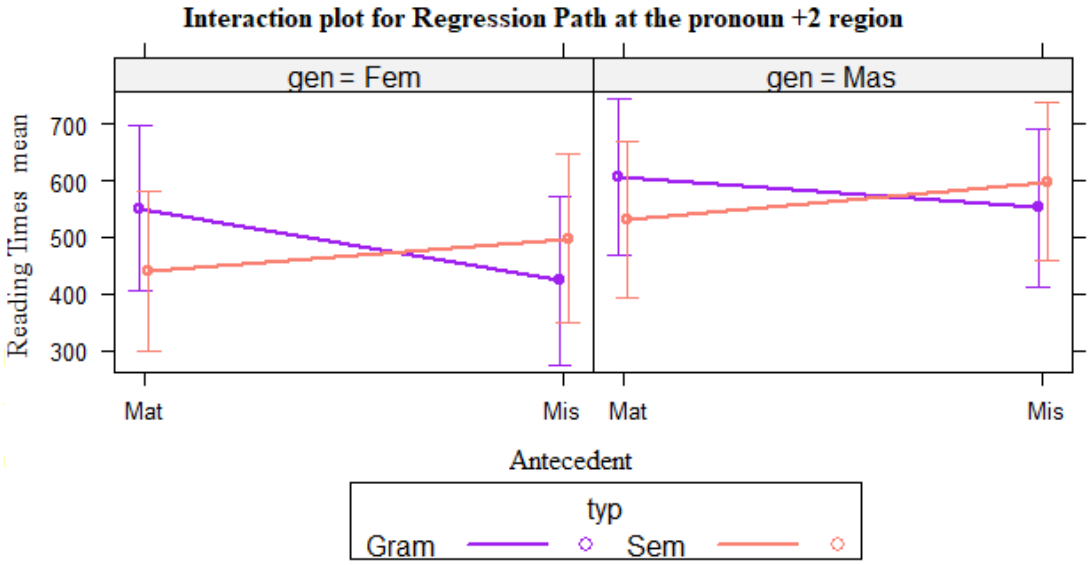

Figure 31: Interaction plot for Regression Path at the pronoun +2 region by antecedent, type of gender, and gender in Experiments 1a and 1b

Tukey HSD tests did not find any statistically significant difference in Figure 31.

Pronoun +3 region

The reading times and standard deviations at the pronoun +3 region in milliseconds are reported in Table 23.

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	796 (1400)	904 (910)	536 (394)	753 (1363)

Mis_Mis	471 (184)	635 (632)	653 (732)	627 (934)
Mat_Mat	591 (1135)	532 (388)	452 (216)	586 (693)
Mat_Mis	506 (338)	805 (997)	695 (1062)	560 (349)

Table 24: Regression Path reading times and standard deviations in parenthesis at the pronoun +3 region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for Regression Path at the pronoun+3 region revealed a statistically significant interaction between antecedent and distractor: F(1, 676648)=3.95, p=0.047. The interaction plot below portraits this interaction.

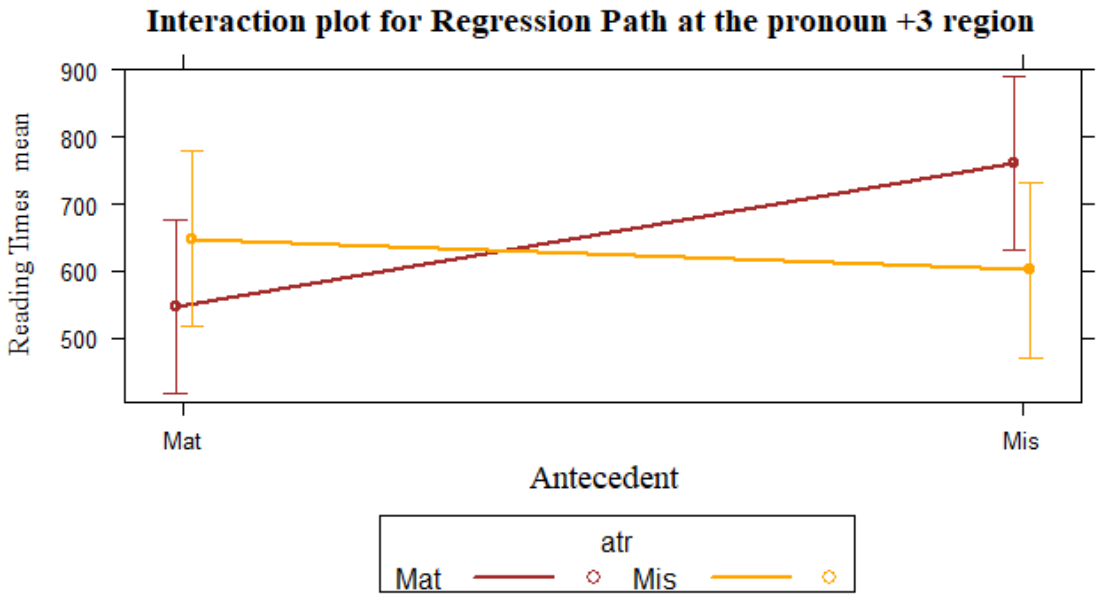

Figure 32: Interaction plot for Regression Path at the pronoun +3 region by antecedent and distractor for Experiments 1a and 1b

Although it was only a trend towards significance, Tukey HSD tests showed reading times at the pronoun +3 region is 213ms longer for mismatching antecedents combined with matching antecedents than for matching antecedents combined with mismatching antecedents (p=0.104).

Discussion

Regression Path, or Go-Part Time, corresponds to the duration of all fixations from first fixation on a region to first moving to the right – this included regressions back to earlier parts of the sentence before moving on. It can reflect either difficulties in integrating information in the context, which can be considered an early effect in processing; or cost of overcoming a difficulty, which can be considered a late effect in processing. At this point both structural and morphological cues are involved in memory retrieval.

Retrievals are faster when the antecedents match the pronouns in gender than when they mismatch, which highlights the importance of structural constraints.

Moreover, since masculine is the default gender, it is more preferably retrieved and weighed more in memory than feminine. Consequently, masculine distractors, matching or mismatching the pronouns, slow down antecedent retrievals due to the fact it causes great similarity-based interference effects and cue confusion.

Second Pass

Figure 33 corresponds to the Second Pass at all regions in Experiments 1a and 1b.

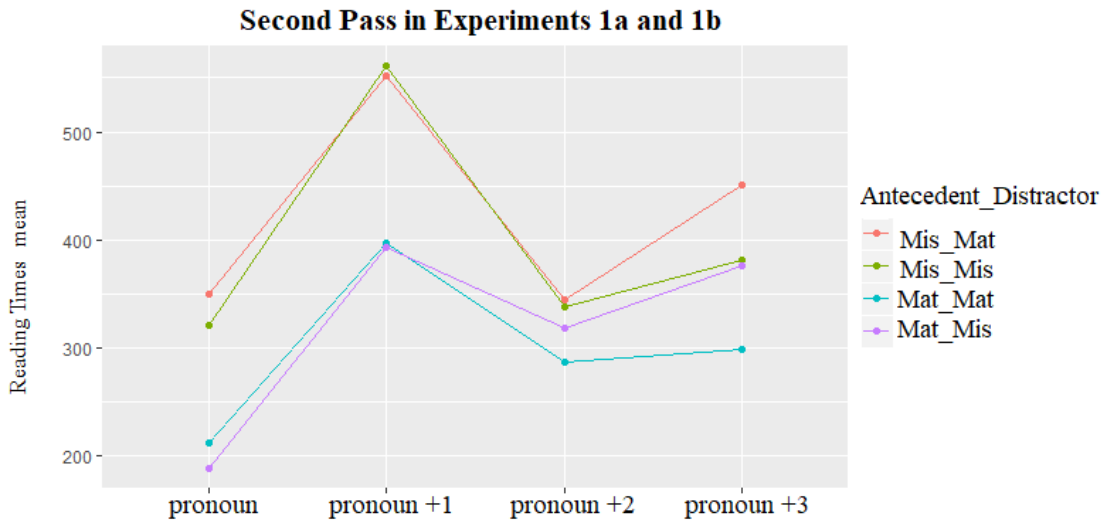

Figure 33: Line chart for Second Pass by regions of interest and conditions in Experiments 1a and 1b

Tukey HSD tests point out that the following contrasts between the conditions are statistically significant: Mis_Mat is 137ms longer than Mat_Mat ($p<0.05$); Mis_Mis is 161ms longer than Mat_Mis ($p<0.05$); Mis_Mis is 108ms longer than Mat_Mat ($p=0.015$); and Mis_Mis is 132ms longer than Mat_Mis ($p=0.001$).

Pronoun region

The reading times and standard deviations at the pronoun region in milliseconds are reported in Table 24. Tukey HSD tests also indicate that reading times at the pronoun region are 122ms faster for matching antecedents than for mismatching antecedents ($p<0.005$). In addition, reading times at the pronoun region are 188ms longer for masculine distractors than for feminine distractors ($p<0.005$).

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	206 (259)	554 (391)	330 (302)	335 (193)
Mis_Mis	274 (317)	433 (349)	182 (236)	406 (325)

Mat_Mat	117 (182)	310 (160)	126 (212)	410 (300)
Mat_Mis	155 (157)	252 (157)	144 (199)	299 (176)

Table 25: Regression Path reading times and standard deviation in parenthesis at the pronoun region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for Second Pass at the pronoun region showed statistically significant main effects of *antecedent*: F(1, 59439)=22.96, p<0.05 and *gender*: F(1, 59439)=34.71, p<0.05. In addition, it was found statistically significant interactions between *antecedent*, *distractor type*, and *distractor gender*: F(1, 59439)=6.45, p=0.011 and between *distractor*, *distractor type of gender*, and *distractor gender*: F(1, 59439)=4.31, p=0.038. Figure 34 and 35 aim at illustrating these interactions.

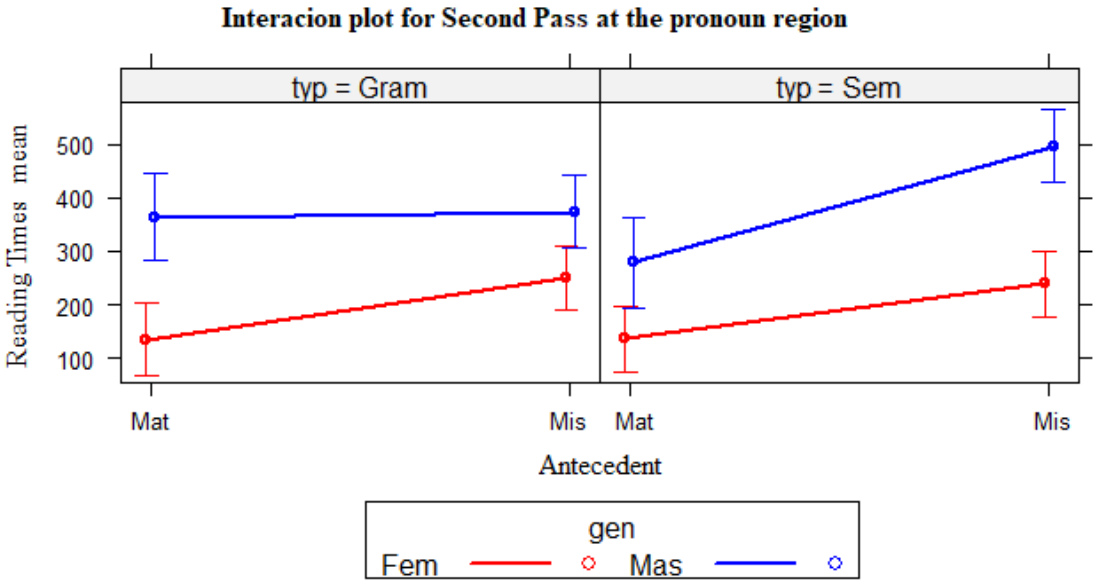

Figure 34: Interaction plot for Second Pass at the pronoun region by antecedent, type of gender, and gender for Experiments 1a and 1b

Tukey HSD tests indicated reading times at the pronoun region in sentences with matching antecedents were 225 ms longer for distractors with masculine grammatical gender than for distractors with feminine gender (p<0.05). Similarly,

reading times in sentences with mismatching antecedents were 255ms longer for distractors with masculine semantic gender than for distractors with feminine semantic gender ($p<0.05$). In addition, matching distractors with masculine semantic gender speed up reading times in 213ms when compared to mismatching distractors ($p=0.002$).

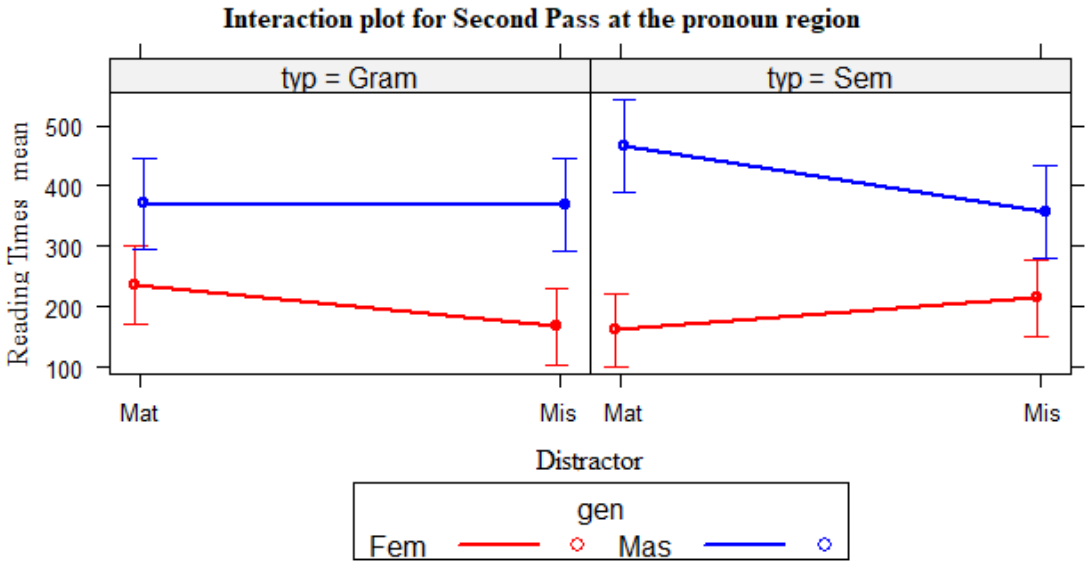

Figure 35: Interaction plot for Second Pass at the pronoun region by distractor, type of gender, and gender in Experiments 1a and 1b

Tukey HSD tests detected reading times at the pronoun region in sentences were 137ms longer for matching distractors with masculine grammatical gender than for matching distractors with feminine grammatical gender in a trend towards significance ($p=0.099$). Similarly, reading times were 200ms longer for mismatching distractors with masculine grammatical gender than for mismatching distractors with feminine grammatical gender ($p=001$).

With respect to semantic gender, reading times at the pronoun region were 282ms longer for matching distractors with masculine semantic gender than for

matching distractors with feminine semantic gender ($p<0.05$). Similarly, reading times were 131ms longer for mismatching distractors with masculine semantic gender than for mismatching distractors with feminine semantic gender ($p=136$) in a trend towards significance.

Pronoun +1 region

The reading times and standard deviations at the pronoun +1 region in milliseconds are reported in Table 26. Tukey HSD tests reveal that matching reading times at the pronoun region +1 region are 142ms faster when antecedents match the pronouns than when they mismatch ($p<0.05$). Moreover, reading times for masculine distractors are 327ms longer than feminine distractors ($p<0.05$).

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	293 (349)	958 (704)	422 (432)	516 (346)
Mis_Mis	430 (462)	774 (510)	275 (319)	687 (569)
Mat_Mat	232 (303)	478 (262)	297 (472)	662 (443)
Mat_Mis	293 (307)	552 (615)	333 (348)	449 (320)

Table 26: Second Pass reading times and standard deviations in parenthesis at the pronoun +1 region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for Second Pass at the pronoun+1 region showed statistically significant main effects of *antecedent*: $F(1, 170688)=14.43$, $p<0.05$ and *distractor gender*: $F(1,170688)=31.02$, $p<0.05$. In addition, it was found statistically significant interactions between *antecedent* and *distractor type of gender*: $F(1, 170688)=5.36$, $p=0.020$, between *antecedent* and *distractor gender*: $F(1, 170688)=4.01$, $p=0.045$, and between *antecedent*, *distractor*,

distractor type of gender, and *distractor gender*: F(1, 170688)=6.58, p=0.010. Figures 36, 37 and 38 clarify these interactions.

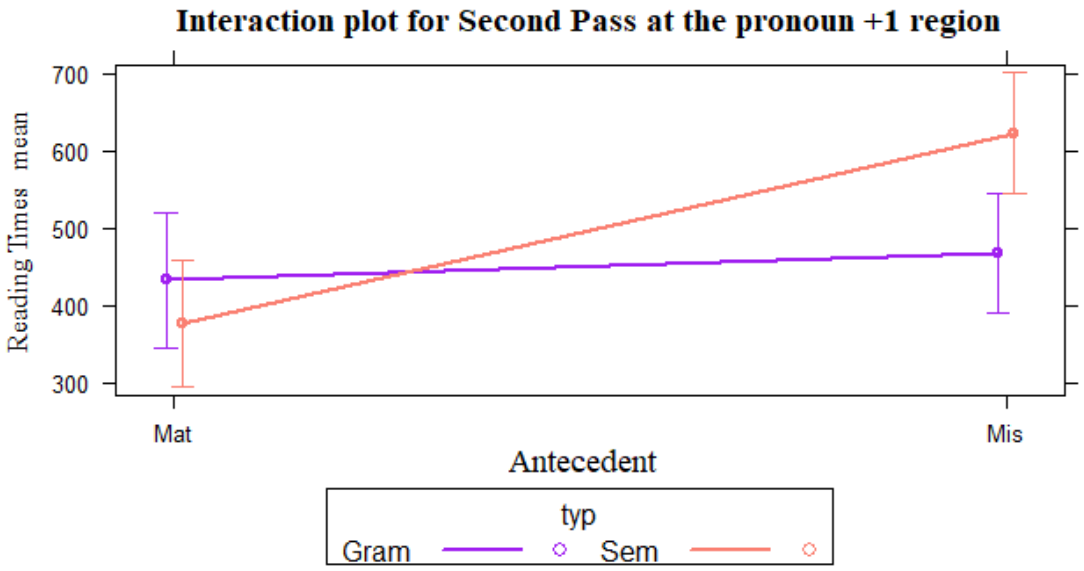

Figure 36: Interaction plot for Second Pass at the pronoun +1 region by antecedent and type of gender in Experiments 1a and 1b

Tukey HSD tests showed reading times at the pronoun +1 region were 139ms longer for sentences with mismatching antecedents followed by distractors with semantic gender when compared to distractors with grammatical gender (p=0.037). Moreover, reading times at the pronoun +1 region for sentences with mismatching antecedents with semantic gender were 230ms longer than for sentences with matching antecedents (p<0.05).

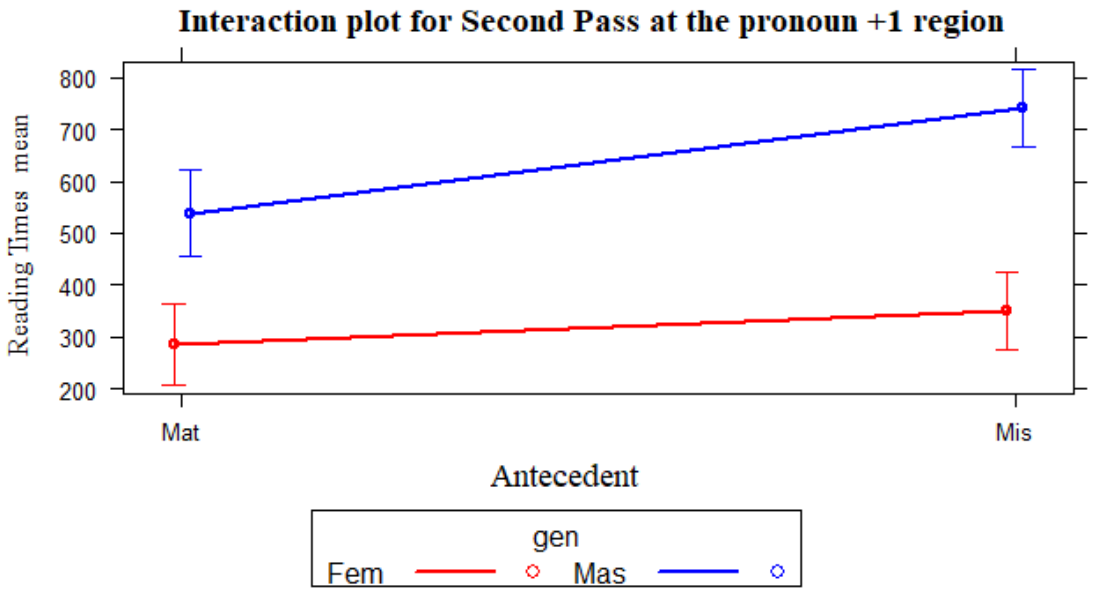

Figure 37: Interaction plot for Second Pass at the pronoun +1 region by antecedent and gender for Experiments 1a and 1b

Tukey HSD tests detected reading times at the pronoun +1 region in sentences with matching antecedents were 255ms longer for masculine distractors than for feminine distractors ($p<0.05$). In sentences with mismatching antecedents, reading times were 387ms longer for masculine distractor than for feminine distractors ($p<0.005$). Furthermore, reading times were 200ms longer for distractors with masculine gender in sentences with mismatching antecedents than in sentences with matching antecedents ($p=0.002$).

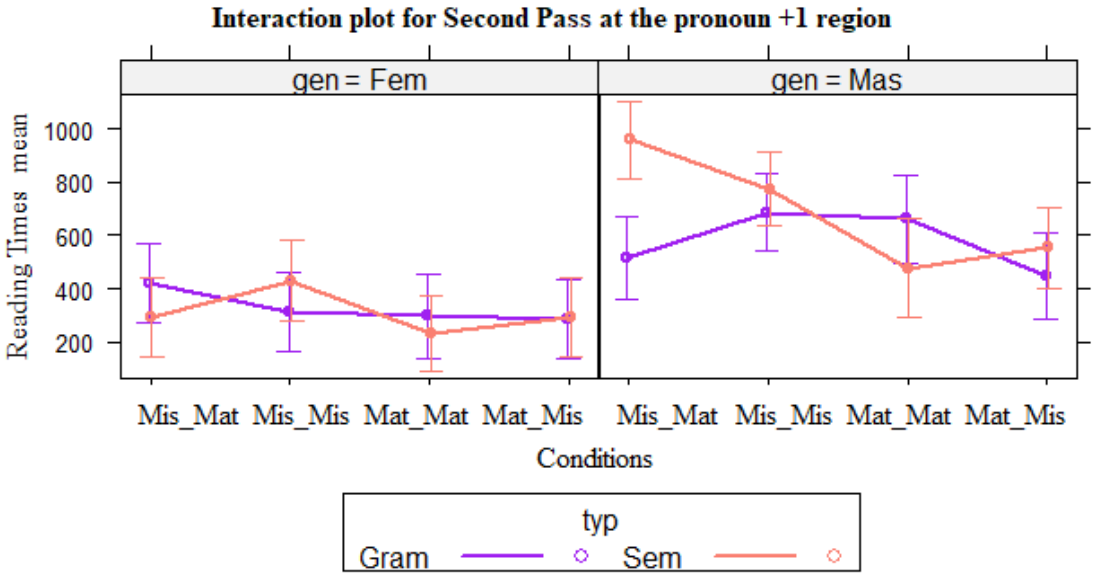

Figure 38: Interaction plot for Second Pass at the pronoun +1 region by conditions, type of gender, and gender in Experiments 1a and 1b

Tukey HSD tests revealed when both antecedents and distractors match the pronouns, reading times at the pronoun +1 region are 364ms longer for distractors with masculine grammatical gender and for distractors with feminine grammatical gender in a trend towards significance (p=0.121). Similarly, when both antecedents and distractors mismatch the pronouns, reading times at the pronoun +1 region are 411ms longer for distractors with masculine grammatical gender than for distractors with feminine grammatical gender (p=0.005).

With respect to semantic gender, when both antecedents and distractors mismatch the pronouns, reading times at the pronoun +1 region are 343ms longer for distractors with masculine semantic gender than for distractors with feminine semantic gender in a trend towards significance (p=0.077).

In addition, in sentences with mismatching antecedents followed by matching antecedents, reading times at the pronoun +1 region are 442ms longer for distractors

with masculine semantic gender than for distractors with masculine grammatical gender (0.004).

Pronoun +2 region

The reading times and standard deviations at the pronoun +2 region in milliseconds are reported in Table 26. It was observed in Tukey HSD tests that antecedent retrievals at the pronoun +2 region are 231ms longer for masculine distractors than for feminine ones ($p<0.005$).

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	195 (299)	567 (441)	233 (242)	413 (373)
Mis_Mis	288 (365)	486 (396)	213 (277)	357 (212)
Mat_Mat	172 (244)	540 (300)	162 (272)	415 (276)
Mat_Mis	294 (298)	478 (281)	231 (262)	369 (196)

Table 27: Second Pass reading times and standard deviations in parenthesis at the pronoun +2 region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for Second Pass at the pronoun +2 region showed a statistically significant main effect of distractor gender: $F(1, 79630)=41.81$, $p<0.05$ and a statistically significant interaction between distractor and distractor gender: $F(1, 79630)=5.52$, $p=0.019$. An interaction plot in Figure 39 illustrates this interaction.

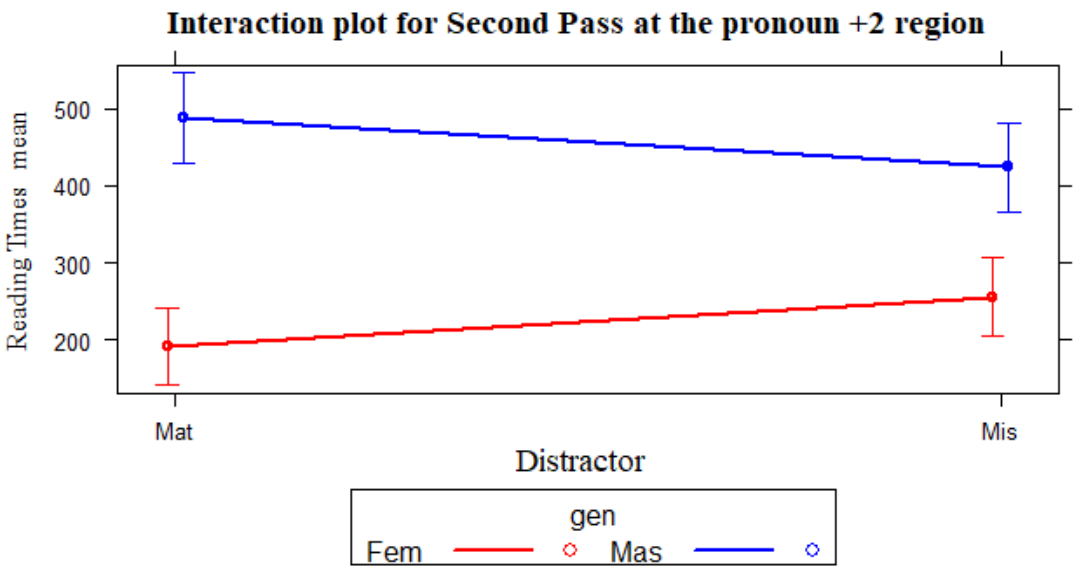

Figure 39: Interaction plot for Second Pass at the pronoun +2 region by distractor and gender in Experiments 1a and 1b

Tukey HSD tests showed reading times at the pronoun +2 region were 295ms longer for matching distractors with masculine gender than for matching distractors with feminine gender ($p<0.05$). Similarly, reading times at the pronoun +2 region were 16ms longer for mismatching distractors with masculine gender than for mismatching distractors with feminine gender ($p<0.05$).

Pronoun +3 region

The reading times and standard deviations at the pronoun +3 region in milliseconds are reported in Table 28. It was detected in Tukey HSD tests that antecedent retrievals at the pronoun +3 region are 66ms faster for antecedents that match the pronouns than for antecedents that mismatch them ($p=0.044$). Moreover, antecedents retrievals are 221ms longer for masculine distractors than for feminine distractors ($p<0.005$).

	Semantic gender		Grammatical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	234 (227)	586 (477)	453 (417)	594 (396)
Mis_Mis	253 (349)	636 (382)	285 (314)	380 (334)
Mat_Mat	249 (301)	347 (226)	185 (352)	446 (347)
Mat_Mis	269 (272)	471 (492)	330 (359)	528 (606)

Table 28: Second Pass reading times and standard deviations in parenthesis at the pronoun +3 region in Experiments 1a and 1b

ANOVAs of the *lmes* of Experiments 1a and 1b for Second Pass at the pronoun +3 region showed statistically significant main effects of antecedent: F(1, 122673)=5.03, p=0.025 and distractor gender: F(1, 122673)=26.05, p<0.05. In addition, it was found statistically significant interactions between antecedent and distractor: F(1, 122673)=7.19, p=0.007, between antecedent, distractor and distractor type of gender: F(1, 122673)=4.02, p=0.045, and between antecedent, distractor type of gender, and distractor gender: F(1, 122673)=5.83, p=0.016. Figures 40, 41 and 42 represent these interactions.

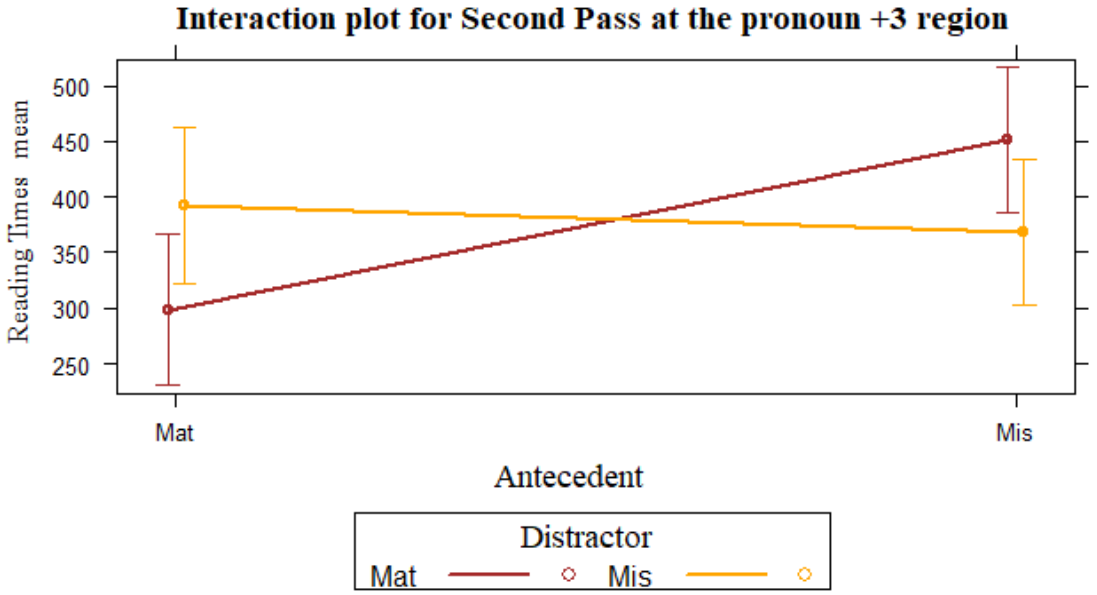

Figure 40: Interaction plot for Second Pass at the pronoun +3 region by antecedent and distractor in Experiments 1a and 1b

Tukey HSD tests revealed reading times at the pronoun +3 region were 145ms longer for mismatching antecedents with matching distractors than for matching antecedents with matching distractors (p=0.008).

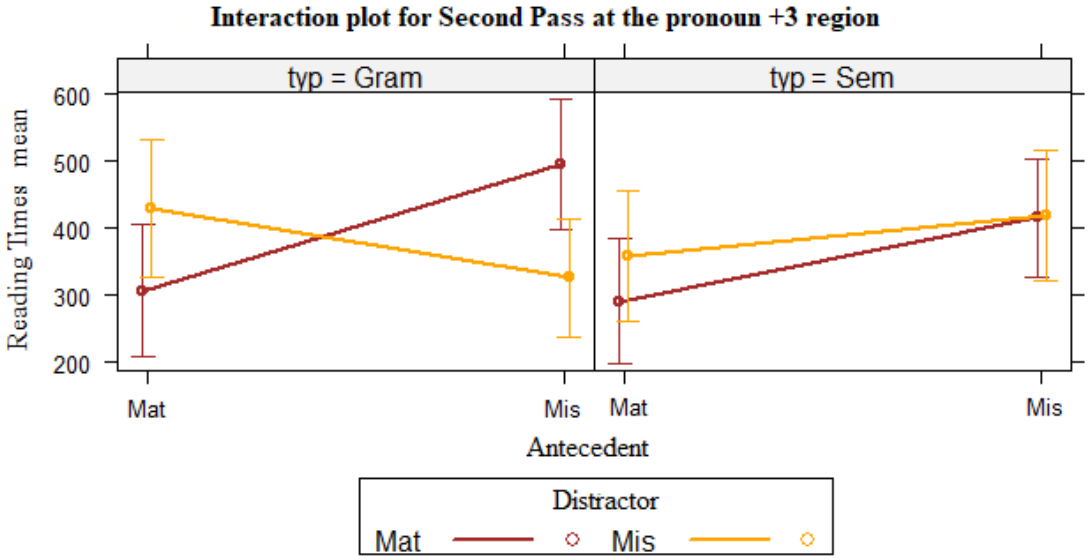

Figure 41: Interaction plot for Second Pass at the pronoun +3 region by antecedent, distractor, and type of gender in Experiments 1a and 1b

Tukey HSD tests showed reading times at the pronoun +3 region were 199ms longer for mismatching antecedents followed by matching distractors with grammatical gender than for matching antecedents with grammatical gender in a trend towards significance (p=0.067).

In addition, reading times at the pronoun +3 region were 174ms faster for mismatching antecedents followed by mismatching distractors with grammatical gender than for matching distractors with grammatical gender in a trend towards significance (p=0.119).

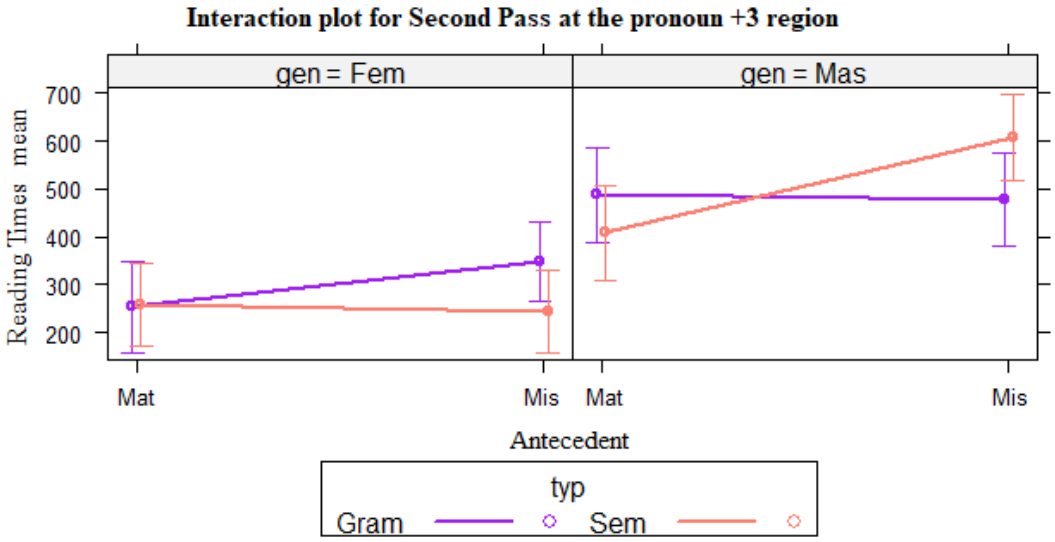

Figure 42: Interaction plot for Second Pass at the pronoun +3 region by antecedent, type of gender, and gender in Experiments 1a and 1b

Tukey HSD tests detected reading times at the pronoun +3 region were 233ms longer for matching antecedents followed by distractors with masculine grammatical gender than for matching antecedents with distractors with feminine grammatical gender ($p=0.019$). Similarly, reading times were 360ms longer for mismatching antecedents followed by distractors with semantic masculine gender than for mismatching antecedents with distractors with feminine semantic gender ($p<0.05$).

In addition, reading times at the pronoun +3 region were 193ms longer for mismatching antecedents followed by distractors with masculine semantic gender than for matching antecedents in a trend towards significance ($p=0.083$).

Discussion

Second Pass corresponds to all the re-fixations in a region of interest, and this is the reason why it is considered a late eye-tracking measure. The results suggest that both structural and morphological cues play a role at this point in processing.

As expected, the influence of binding structural constraints was responsible for faster antecedent retrievals in sentences with matching antecedents than in sentences with mismatching antecedents.

On the other hand, as expected, the influence of gender cues was responsible for longer antecedent retrievals in sentences with masculine distractors than in sentences with feminine distractors for both semantic and grammatical gender, and in sentences with both matching or mismatching antecedents. This result was expected since masculine rather than feminine is more preferably retrieved in memory. In other words, masculine weighed more in memory than feminine so that masculine distractors have increased prominence, causing inhibitory similarity-based interference effects (Lewis & Vasishth, 2005; Engelmann et al., 2015).

Importantly, the presence of matching distractors with masculine semantic gender facilitated antecedent retrievals when compared to mismatching distractors. This facilitatory effect, called intrusion, means the distractors are being misretrieved as if they were the antecedents.

Moreover, following the predictions, when antecedents mismatch the pronouns, distractors with semantic gender were responsible for slower antecedents retrievals than distractors with grammatical gender. A reason for that is semantic gender is preferably retrieved rather than grammatical gender. Semantic gender may weigh more in memory due to the fact it has both syntactic and conceptual genders. Since semantic gender is prominent, it causes cue confusion in memory due to the fact they compete with the mismatching antecedents. Thus, slower antecedent retrievals occur.

4.4.2.6 Discussion of Experiments 1a and 1b

The results of Experiments 1a and 1b suggested both pronominal structural cues and gender morphological cues are equally taken into account throughout antecedent memory retrieval and pronoun processing. The results showed the structural constraints of Principle B does not work as an initial filter blocking the interference of structurally unacceptable candidates at early processing stages; on the contrary, structurally unacceptable candidates seem to strongly affect antecedent retrieval since the very first processing phase (First Pass), even when there is a structurally acceptable antecedent available in the sentence that matches the pronoun.

As expected, although structural constraints did not filter the antecedent candidates, it greatly contributed to antecedent retrieval. Pronoun resolution was facilitated by the presence of a structurally acceptable antecedent that matched the pronouns in gender rather than the presence of a structurally acceptable antecedent that mismatched the pronouns. Therefore, what matters for pronoun processing is not only the presence of morphological matching candidates in the sentence, but also the presence of structurally acceptable candidates that morphologically match the pronouns. The feature [+ACCESSIBLE] conveyed by structurally acceptable antecedents (Kush et al., 2015) is as important as the morphological cues. However, sometimes, distractors, which are candidates that do not have the [+ACCESSIBLE] feature, can be misretrieved by memory as if they were the antecedents.

As predicted semantic gender seemed to weigh more in memory than grammatical gender since structurally unacceptable candidates carrying semantic gender caused more similarity-interference effects (Lewis & Vasishth, 2005) than grammatical gender. An explanation for that lies in the fact semantic gender is redundant, that is, it carries both syntactic and semantic gender, while grammatical

gender only carries syntactic gender (Vigliocco & Franck, 1999). Thus semantic gender is more prominent than grammatical gender (Engelmann et al., 2015). However, it should be highlighted distractors with semantic gender can either cause inhibitory, which happens in the majority of cases, or facilitatory effects. In the first case, distractors with semantic gender compete with antecedents in memory retrieval, which slows down retrievals; in the second case, distractors with semantic gender are misretrieved, which speeds up retrievals.

As already predicted, masculine gender seemed to weigh more in memory than feminine gender once masculine structurally unacceptable candidates caused more similarity-based interference effects than feminine structurally unacceptable candidates (Lewis & Vasishth, 2005). A reason for that is masculine is the default gender in Portuguese, that is, it is a neutral generic gender that might embody both masculine and feminine referent representations (Corbett, 1991; Casado et al., 2017). Masculine gender seems to be more prominent in memory than feminine gender (Engelmann et al., 2015). However, masculine distractors can also either cause an inhibitory, which happens in the majority of cases, or a facilitatory effect, that is, distractors with masculine gender can either slow down antecedent retrievals as they compete with antecedents in memory or speeds up retrievals as they can be misretrieved in memory as if they were the antecedents.

Curiously, there were not any major differences among the regions of interest analyzed in Experiments 1a and 1b. It was expected that some effects generated by the time readers got to the pronoun region would be eventually detected in the spillover regions. However, this was not the case. In fact, it seems the 4 regions of interest, namely, the *pronoun region*, the *pronoun +1 region*, the *pronoun +2 region*, and the *pronoun +3 region*, reflected the same kinds of results. This means all the

effects originated at the pronoun region are taken over across the spillover regions. Similarly, there were no major differences among the eye-tracking measures; both early and late measures captured the same kinds of results. Taken together, these two pieces of evidence corroborate to the idea that antecedent retrieval in Brazilian Portuguese is a solid and constant process that simultaneously takes both structural and morphological cues into account throughout processing, with no exclusive early or late effects.

In order to better understand the role of morphological cues in antecedent retrieval, a comparison between definitional gender and stereotypical gender would be fruitful in disentangling respectively which gender cues would weigh more in memory, lexical gender cues or world-knowledge based gender cues.

4.4.3 Experiments 2a and 2b

Experiments 2a and 2b aimed at investigating the differences between definitional gender and stereotypical gender during antecedent retrieval in Brazilian Portuguese. By comparing these two types of genders, it would be possible to verify whether definitional gender would weigh more (see section 2.6.5) in memory than stereotypical gender. Antecedents with definitional gender would be more preferable candidates than antecedents with stereotypical gender. A reason for that lies in the fact that definitional gender is lexically specified, while stereotypical gender is a result of probabilistic world knowledge inferences (Carreiras et al., 1996; Kneiner et al., 2008; Oakhill et al., 2008).

Since definitional gender would be more preferable than stereotypical gender, pronoun processing with structurally unacceptable candidates carrying definitional

gender would be more costly than pronoun processing with structurally unacceptable candidates carrying stereotypical gender. Structurally unacceptable antecedents carrying definitional gender would cause larger interference effects, because there would be more competition between them and structurally acceptable antecedents, and, consequently, antecedent retrievals would be slower.

Moreover, since masculine is the default gender (Corbett, 1991; Casado et al., 2017), it is expected masculine structurally unacceptable candidates would be responsible for slower antecedent retrievals. They would cause greater interference effects than feminine since they are more preferable to be retrieved.

Both Experiments 2a and 2b tested structurally unacceptable antecedents with definitional and stereotypical genders; however, the former tested those types of gender in the feminine and the latter in the masculine as illustrated in the table below.

Eye-tracking experiment	Type of gender	Structurally unacceptable antecedent candidate examples
Experiment 2a	Feminine definitional gender	*mulher* (woman)
	Feminine stereotypical gender	*recepcionista* (receptionist)
Experiment 2b	Masculine definitional gender	homem (man)
	Feminine stereotypical gender	surfista (surfist)

Table 29: Experiments 2a and 2b

4.4.3.1 Participants

32 native speakers of Brazilian Portuguese (21 female and 11 male, with a mean age of 22 years) participated in Experiment 2a; and 36 (24 female and 12 male, average of

age of 22 years) participated in Experiment 2b. The selection and compensation for the participants' work were the same of Experiments 1a and 1b.

4.4.3.2 Materials and design

We used the same materials and design of Experiments 1a and 1b, except for the gender of the attractors, which in this experiment can be definitional or stereotypical. A sample of the materials used in Experiment 2a can be seen in Tables 29 and 30, while a sample of the materials used in Experiment 2b can be seen in Tables 31 and 32. Brackets delimit the regions of interest.

(86) Distractor with feminine definitional gender:

	Antecedent mismatch	**Antecedent match**
Attractor mismatch	A enfermeira conhecia a mulher que matou [**ele**] [brutalmente] [na frente] [da casa da] família. (*The nurse$_{[fem]}$ knew the woman who brutally killed **him** in front of the family's house.)*	O enfermeiro conhecia a mulher que matou [**ele**] [brutalmente] [na frente] [da casa da] família. (*The nurse$_{[masc]}$ knew the woman who brutally killed **him** in front of the family's house.)*
Attractor match	O enfermeiro conhecia a mulher que matou [**ela**] [brutalmente] [na frente] [da casa da] família. (*The nurse$_{[masc]}$ knew the woman who brutally killed **her** in front of the family's house.)*	A enfermeira conhecia a mulher que matou [**ela**] [brutalmente] [na frente] [da casa da] família. (*The nurse$_{[fem]}$ knew the woman who brutally killed **her** in front of the family's house.)*

Table 30: Sample of materials for distractors with feminine definitional gender used in Experiment 2a by regions of interest

(87) Distractor with feminine stereotypical gender:

	Antecedent mismatch	Antecedent match
Attractor mismatch	A bibliotecária seguiu a recepcionista que guiou [**ele**] [brevemente] [através] [do corredor] do grande gabinete real. (*The librarian$_{[fem]}$ followed the receptionist$_{[fem]}$ who briefly guided **him** through the large royal office.)*	O bibliotecário seguiu a recepcionista que guiou [**ele**] [brevemente] [através] [do corredor] do grande gabinete real. (*The librarian$_{[masc]}$ followed the receptionist$_{[fem]}$ who briefly guided **him** through the large royal office.)*
Attractor match	O bibliotecário seguiu a recepcionista que guiou [**ela**] [brevemente] [através] [do corredor] do grande gabinete real. (*The librarian$_{[masc]}$ followed the receptionist$_{[fem]}$ who briefly guided **her** through the large royal office.)*	A bibliotecária seguiu a recepcionista que guiou [**ela**] [brevemente] [através] [do corredor] do grande gabinete real. (*The librarian$_{[fem]}$ followed the receptionist$_{[fem]}$ who briefly guided **her** through the large royal office.)*

Table 31: Sample of the materials for distractors with feminine stereotypical gender used in Experiments 2a by regions of interest

(88) Distractor with masculine definitional gender:

	Antecedent mismatch	Antecedent match
Attractor mismatch	O advogado hostilizou o rei que tratou tratou [**ela**] [rudemente] [na frente] [de alguns] convidados na festa. (*The lawyer$_{[masc]}$ antagonized the king who treated **her** rudely in front of some guests at the party.)*	A advogada hostilizou o rei que tratou tratou [**ele**] [rudemente] [na frente] [de alguns] convidados na festa. (*The lawyer$_{[fem]}$ antagonized the king who treated **him** rudely in front of some guests at the party.)*
Attractor match	A advogada hostilizou o rei que tratou tratou [**ele**] [rudemente] [na frente] [de alguns] convidados na	O advogado hostilizou o rei que tratou tratou [**ele**] [rudemente] [na frente] [de alguns] convidados na

	festa. (*The lawyer$_{[fem]}$ antagonized the king who treated **him** rudely in front of some guests at the party.)*	festa. (*The lawyer$_{[masc]}$ antagonized the king who treated **him** rudely in front of some guests at the party.)*

Table 32: Sample of the materials for distractors with masculine definitional gender used in Experiment 2b by regions of interest

(89) Distractor with masculine stereotypical gender:

	Antecedent mismatch	**Antecedent match**
Attractor mismatch	A cabeleireira detestou o piloto de corrida que olhou [**ela**] [lentamente] [dos pés] [a cabeça] antes da entrevista no camarim. (*The hair dresser$_{[fem]}$ disliked the race car driver$_{[masc]}$ who slowly stared at **her** in the dressing room before the interview.)*	O cabeleireiro detestou o piloto de corrida que olhou [**ela**] [lentamente] [dos pés] [a cabeça] antes da entrevista no camarim. (*The hair dresser$_{[masc]}$ disliked the race car driver$_{[masc]}$ who slowly stared at **her** in the dressing room before the interview.)*
Attractor match	A cabeleireira detestou o piloto de corrida que olhou [**ele**] [lentamente] [dos pés] [a cabeça] antes da entrevista no camarim. (*The hair dresser$_{[fem]}$ disliked the race car driver$_{[masc]}$ who slowly stared at **him** in the dressing room before the interview.)*	O cabeleireiro detestou o piloto de corrida que olhou [**ele**] [lentamente] [dos pés] [a cabeça] antes da entrevista no camarim. (*The hair dresser$_{[masc]}$ disliked the race car driver$_{[masc]}$ who slowly stared at **him** in the dressing room before the interview.)*

Table 33: Sample of the materials for distractors with masculine stereotypical gender used in Experiment 2b by regions of interest

4.4.3.3 Procedure

The same procedure used in Experiments 1a and 1b.

4.4.3.4 Analysis

The same analysis of Experiments 1a and 1b. Some experimental trials had to be excluded due to eye blinks and long saccades at the regions of interest (22% in Experiment 2a, and 15% in Experiment 2b).

4.4.3.5 Results

It should be mentioned that the participants answered the comprehension questions with an average of accuracy of 90% in Experiment 2a and 89% in Experiment 2b, which means that the participants were paying attention to the task and reading the sentences properly.

Results for each of the eye-tracking measures will be reported and discussed below according to the regions of interest investigated.

First Fixation

For space reasons, only the statistically significant results will be reported in this section.

Figure 43 corresponds to the First Fixation at all regions of interest in Experiments 2a and 2b.

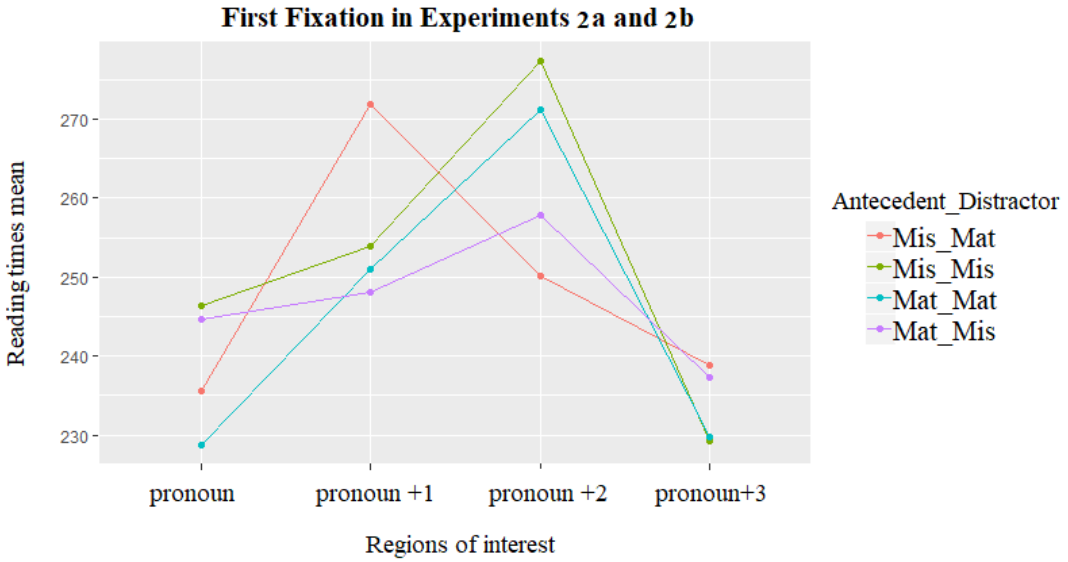

Figure 43: Line chart for First Fixation by regions of interest and conditions in Experiments 2a and 2b

ANOVA tests found statistically significant results for all regions, except for the pronoun +1 region.

Pronoun region

The reading times and standard deviations at the pronoun region in milliseconds are reported in Table 32. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	294 (111)	224 (61)	245 (67)	198 (58)
Mis_Mis	276 (131)	251 (103)	227 (64)	226 (101)
Mat_Mat	257 (107)	207 (75)	197 (86)	252 (101)
Mat_Mis	247 (64)	246 (94)	248 (123)	238 (91)

Table 34: First Fixation reading times and standard deviations in parenthesis at the pronoun region in Experiments 2a and 2b

ANOVAs of the *lmes* for First Fixation at the pronoun region revealed a statistically significant main effect of *distractor gender type*: F(1, 8007)=5.02, p=0.025.

Tukey HSD tests showed reading times at the pronoun region are 18ms faster for sentences with masculine distractors than for sentences with feminine distractors with a trend towards significance (p=0.055).

Pronoun +2 region

The reading times and standard deviations at the pronoun region in milliseconds are reported in Table 33. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	255 (98)	247 (99)	249 (81)	249 (92)
Mis_Mis	261 (75)	268 (87)	252 (71)	315 (183)
Mat_Mat	274 (106)	257 (99)	241 (89)	302 (123)
Mat_Mis	278 (107)	257 (87)	258 (81)	243 (79)

Table 35: First Fixation reading times and standard deviations in parenthesis at the pronoun +2 region in Experiments 2a and 2b

ANOVAs of the *lmes* for First Fixation at the pronoun +2 region revealed statistically significant interactions between *distractor type of gender* and *distractor gender*: F(1, 9025)=6.38, p=0.011 and between *antecedent*, *distractor*, and *distractor gender*: F(1, 9025)=6.27, p=0.012. Interaction plots below illustrate these interactions.

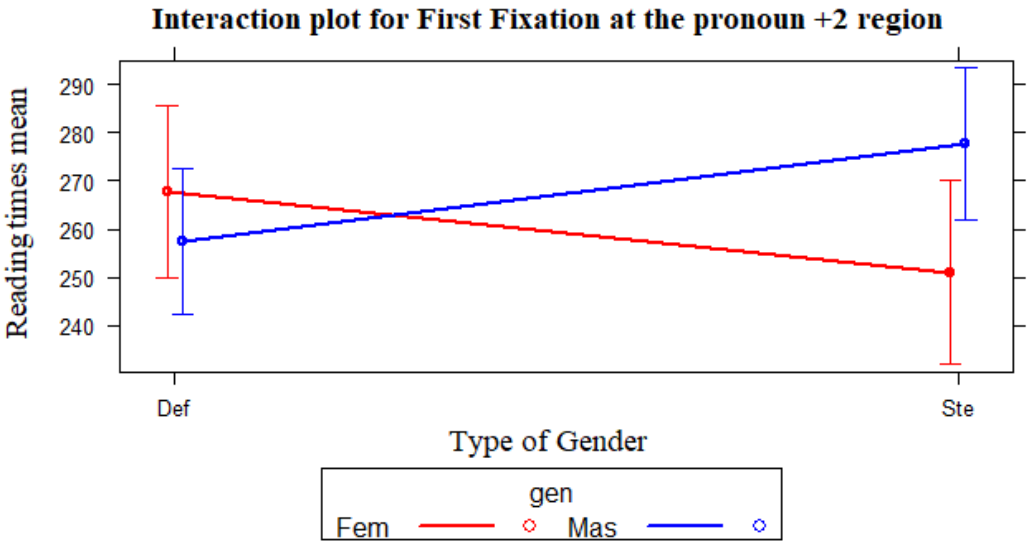

Figure 44: Interaction plot for First Fixation at the pronoun +2 region by type of gender in Experiments 2a and 2b

Tukey HSD tests indicated reading times at the pronoun +2 region were 26ms longer for distractors with masculine stereotypical gender than for distractors with feminine stereotypical gender in a trend towards significance (p=0.156).

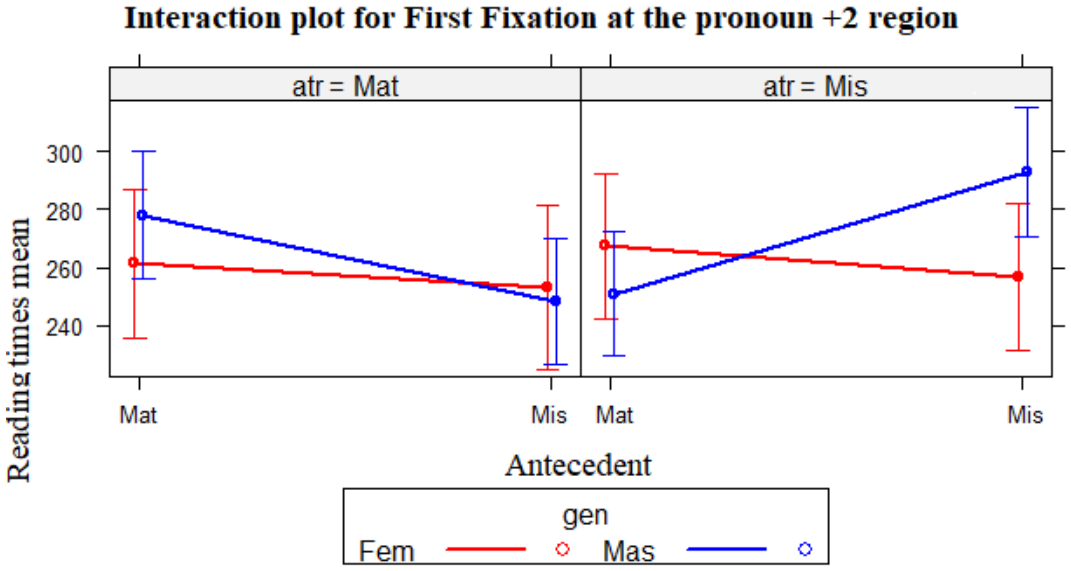

Figure 45: Interaction plot for First Fixation at the pronoun +2 region by antecedent, distractor, and gender in Experiments 2a and 2b

Tukey HSD tests revealed reading times at the pronoun +2 region were 44ms longer in sentences in which both antecedents and masculine distractors mismatch the pronouns than in sentences with a mismatching antecedent followed by a matching masculine distractor in a trend towards significance (p=0.089).

Pronoun +3 region

The reading times and standard deviations at the pronoun region in milliseconds are reported in Table 46. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	229 (91)	237 (80)	248 (69)	241 (86)
Mis_Mis	259 (117)	217 (71)	221 (89)	223 (53)
Mat_Mat	237 (73)	221 (69)	243 (80)	222 (65)
Mat_Mis	250 (155)	250 (111)	208 (84)	238 (71)

Figure 46: First Fixation reading times and standard deviations in parenthesis at the pronoun +3 region in Experiments 2a and 2b

ANOVAs of the *lmes* for First Fixation at the pronoun +3 region revealed statistically significant interactions between *distractor* and *distractor type of gender*: F(1, 6797)=3.89, p=0.048. The interaction plot below illustrates this interaction.

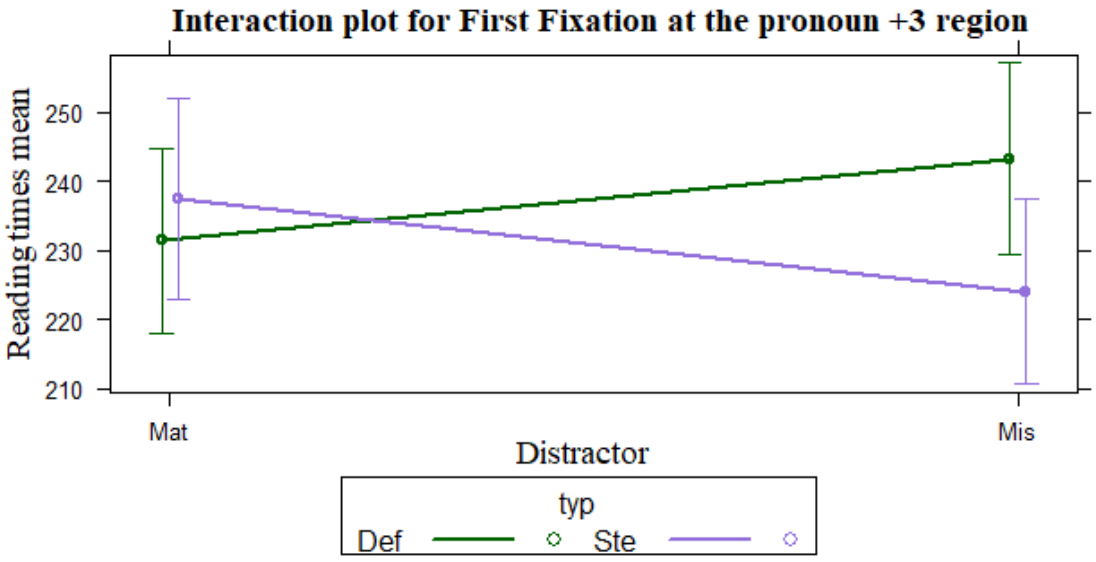

Figure 47: Interaction plot for First Fixation at the pronoun +2 region by antecedent, distractor, and gender in Experiments 2a and 2b

Tukey HSD tests did not show any statistically significant difference in Figure 47.

Discussion

The results for First Fixation indicated both structural and morphological cues are equally involved in early pronoun processing stages.

It seems distractors were misretrieved as if they were antecedents in the absence of a structurally acceptable antecedent that matches the pronoun.

In addition, it appears masculine distractors facilitated reading times when compared to feminine distractors due to the fact that masculine weighed more in memory than feminine. Masculine gender is the default gender in Portuguese; consequently, masculine distractors are preferably retrieved and more prominent candidates. Thus they can sometimes be misretrieved by memory as if they were the antecedents, speeding up retrieval times. However, not always masculine distractors

have a facilitatory effect. For example, distractors with masculine stereotypical slowed down retrievals when compared to distractors with feminine stereotypical gender. In this latter case, what happened was that masculine stereotypical distractors might compete with antecedents in memory, causing an inhibitory effect.

First Pass

Figure 48 corresponds to the Firs Pass reading times at all regions of interest in Experiments 2a and 2b.

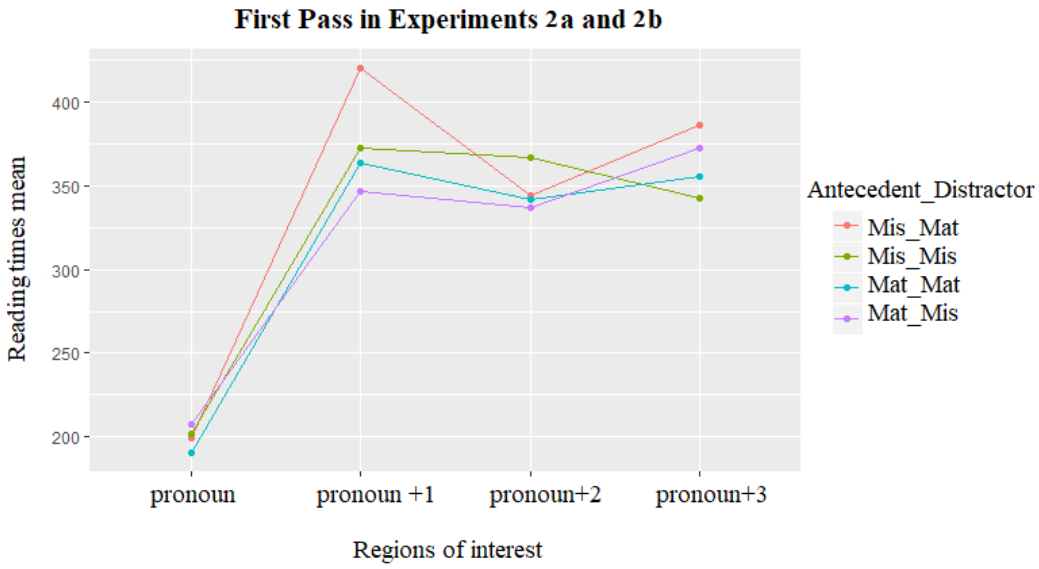

Figure 48: Line chart for First Pass by regions of interest and conditions in Experiments 2a and 2b

ANOVA tests found statistically significant results for pronoun region and pronoun +2 region.

Pronoun region

The reading times and standard deviations at the pronoun region in milliseconds are reported in Table 34. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	174 (182)	240 (93)	151 (155)	225 (81)
Mis_Mis	191 (209)	285 (117)	97 (121)	249 (123)
Mat_Mat	141 (165)	238 (113)	126 (118)	268 (107)
Mat_Mis	155 (143)	246 (94)	163 (163)	267 (131)

Table 36: First Pass reading times and standard deviations in parenthesis at the pronoun region in Experiments 2a and 2b

ANOVAs of the *lmes* for First Fixation at the pronoun region revealed a statistically significant main effect of *distractor gender type*: $F(1,17080)=37.62$, $p<0.05$. Tukey HSD tests shows reading times at the pronoun region is 103ms longer for masculine distractors than for feminine ones ($p<0.05$).

Pronoun +2 region

The reading times and standard deviations at the pronoun +2 region in milliseconds are reported in Table 35. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	337 (201)	370 (166)	286 (177)	356 (116)
Mis_Mis	345 (115)	414 (176)	291 (195)	400 (235)
Mat_Mat	354 (241)	326 (142)	308 (219)	370 (140)
Mat_Mis	358 (157)	345 (108)	321 (196)	325 (116)

Table 37: First Pass reading times and standard deviations in parenthesis at the pronoun +2 region in Experiments 2a and 2b

ANOVAs of the *lmes* for First Fixation at the pronoun +2 region revealed a statistically significant interaction between *antecedent* and *distractor gender*: F(1,24126)=6.64, p=0.010. The interaction plot below clarifies this interaction.

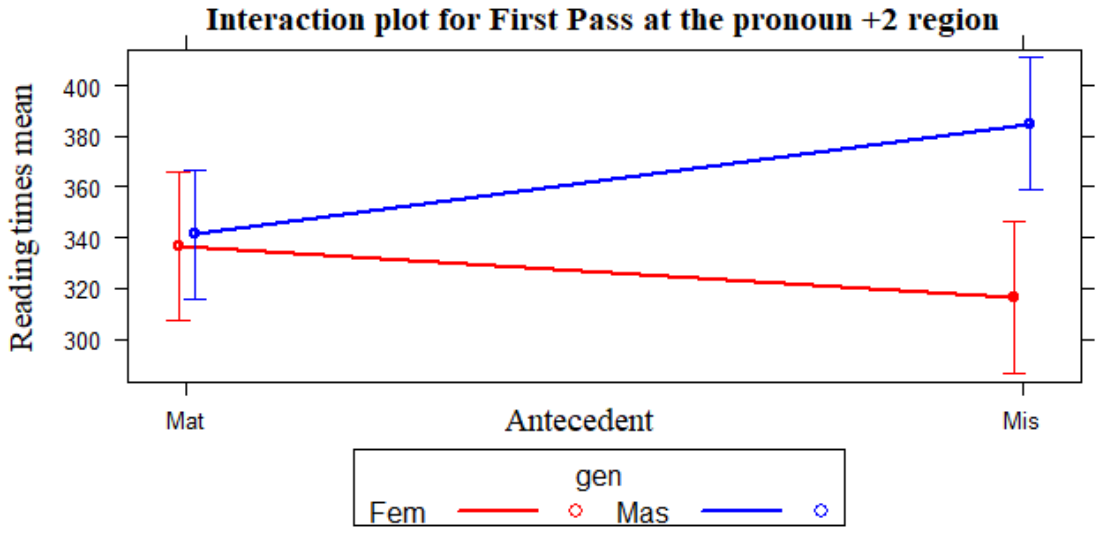

Figure 49: Interaction plot for First Pass at the pronoun +2 region by antecedent and gender in Experiments 2a and 2b

Tukey HSD tests revealed reading times at the pronoun +2 region were 44ms longer in sentences with a mismatching antecedent followed by a masculine distractor than in sentences with a matching antecedent followed by a masculine distractor in a trend towards significance (p=0.074).

Moreover, reading times at the pronoun +2 region were 69ms longer in sentences with mismatching antecedents followed by masculine distractors than in sentences with mismatching antecedents followed by feminine distractors (p=0.003).

Discussion

The results showed both structural constraints and morphological cues play a role in First Pass. Sentences in which the antecedents agreed with the pronouns in gender had faster antecedent retrievals than sentences in which the antecedents disagreed with the pronouns. Additionally, masculine gender seemed to be preferably retrieved in memory than feminine gender; therefore, masculine distractors caused more similarity-based interference effects, that is, greater inhibitory effects, than feminine gender. Masculine is the default gender, which might justify the fact masculine distractors are prominent candidates than feminine distractors.

Regression Path

Figure 50 corresponds to the Regression Path reading times at all regions of interest in Experiments 2a and 2b.

Tukey HSD tests detected reading times at pronoun region for condition Mis_Mis were 136ms longer than for condition Mat_Mat in a trend towards statistical significance ($p=0.169$). Moreover, reading times at the pronoun +1 region for Mis_Mat condition were 370ms longer than in Mat_Mat condition ($p<0.05$); Mis_Mat condition were 432ms longer than in Mat_Mis condition ($p<0.05$); Mis_Mis condition were 345ms longer than Mat_Mat condition ($p<0.05$); Mat_Mat were 405ms longer than Mis_Mis condition ($p<0.05$). In addition, reading times at the pronoun +2 region were 155ms faster for Mis_Mat condition than for Mat_Mis condition ($p=0.032$) and condition Mis_Mis was 122ms faster than condition Mat_Mis in a trend towards significance ($p=0.133$).

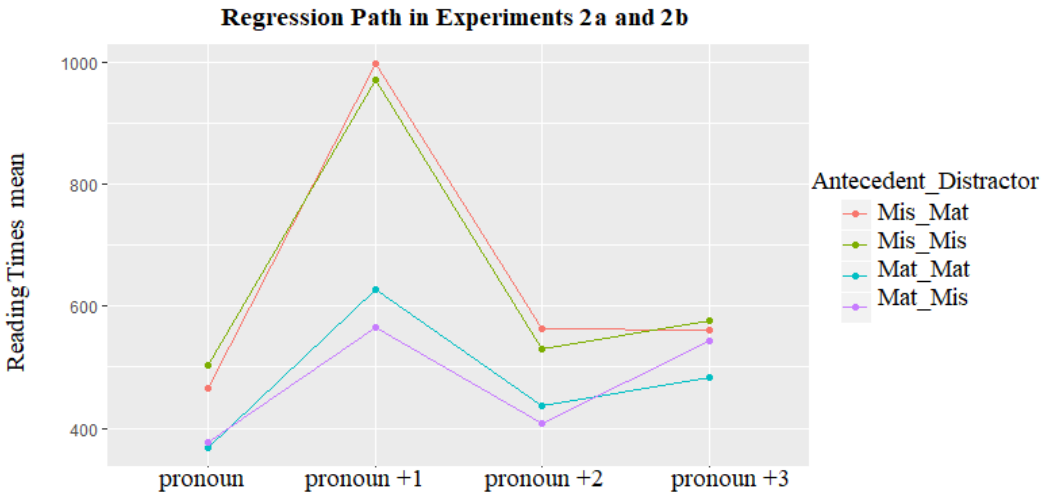

Figure 50: Line chart for Regression Path by regions of interest and conditions in Experiments 2a and 2b

ANOVA tests found statistically significant results for all regions of interest.

Pronoun region

The reading times and standard deviations at the pronoun region in milliseconds are reported in Table 36. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	554 (673)	361 (259)	631 (719)	428 (487)
Mis_Mis	516 (794)	428 (370)	393 (401)	676 (631)
Mat_Mat	276 (126)	358 (491)	254 (128)	416 (324)
Mat_Mis	535 (733)	266 (129)	337 (323)	434 (309)

Table 38: Regression Path reading times and standard deviations in parenthesis at the pronoun region in Experiments 2a and 2b

ANOVAs of the *lmes* for Regression Path at the pronoun region revealed a statistically significant main effect of *antecedent*: F(1, 196687)=7.50, p=0.006 and a statically significant interaction between *antecedent*, *distractor*, and *distractor gender*: F(1, 196687)=9.31, p=0.002.

Tukey tests revealed reading times at the pronoun region were 111ms longer in sentences with mismatching antecedents than in sentences with matching antecedents (p=0.016). The interaction plot below illustrates the interaction found out for pronoun region.

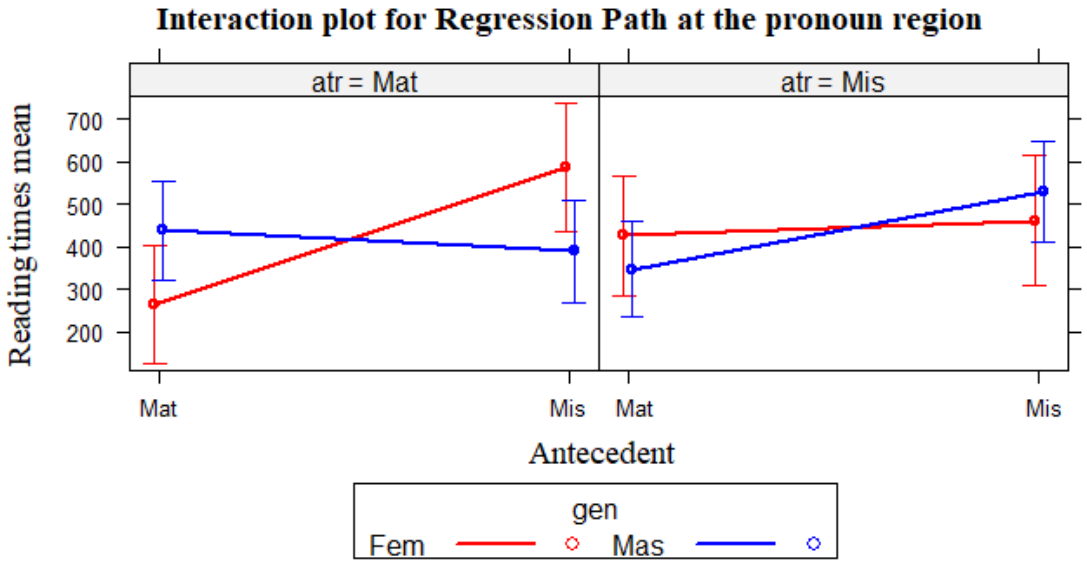

Figure 51: Interaction plot for Regression Path at the pronoun by antecedent, distractor, and gender in Experiments 2a and 2b

Tukey HSD tests showed reading times at the pronoun region were 320ms longer for sentences with mismatching antecedents followed by matching feminine distractors than for sentences with matching antecedents followed by matching feminine distractors (p=0.046).

Pronoun +1 region

The reading times and standard deviations at the pronoun +1 region in milliseconds are reported in Table 37. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	1023 (1098)	361 (926)	1145 (1289)	863 (915)
Mis_Mis	1014 (1056)	428 (808)	907 (872)	938 (1094)
Mat_Mat	674 (1122)	358 (560)	494 (242)	497 (391)
Mat_Mis	625 (576)	266 (407)	457 (299)	511 (471)

Table 39: Regression Path reading times and standard deviations in parenthesis at the pronoun +1 region in Experiments 2a and 2b

ANOVAs of the *lmes* for Regression Path at the pronoun +1 region revealed a statistically significant main effect of *antecedent*: $F(1, 620119)=39$, $p<0.05$.

Tukey tests revealed reading times at the pronoun region were 388ms longer in sentences with mismatching antecedents than in sentences with matching antecedents ($p<0.05$).

Pronoun +2 region

The reading times and standard deviations at the pronoun +2 region in milliseconds are reported in Table 38. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	603 (826)	550 (673)	708 (1123)	455 (434)
Mis_Mis	504 (428)	589 (610)	451 (354)	553 (579)
Mat_Mat	433 (367)	470 (325)	410 (275)	418 (189)

Mat_Mis	483 (448)	420 (266)	342 (213)	392 (210)

Table 40: Regression Path reading times and standard deviations in parenthesis at the pronoun +2 region in Experiments 2a and 2b

ANOVAs of the *lmes* for Regression Path at the pronoun +2 region revealed a statistically significant main effect of *antecedent*: F(1, 242000)=10.77, p=0.001.

Tukey tests revealed reading times at the pronoun region were 124ms longer in sentences with mismatching antecedents than in sentences with matching antecedents (p=0.002).

Pronoun +3 region

The reading times and standard deviations at the pronoun +3 region in milliseconds are reported in Table 39. Tukey HSD tests did not show any statistically significant pairwise comparisons across the different conditions in the experiment.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	526 (427)	782 (1199)	425 (263)	436 (279)
Mis_Mis	798 (979)	594 (814)	491 (783)	382 (172)
Mat_Mat	510 (411)	630 (718)	357 (139)	396 (241)
Mat_Mis	767 (953)	503 (442)	561 (864)	428 (235)

Table 41: Regression Path reading times and standard deviations in parenthesis at the pronoun +3 region in Experiments 2a and 2b

ANOVAs of the *lmes* for Regression Path at the pronoun +3 region revealed a statistically significant main effect of *distractor type of gender*: F(1, 405460)=10.11, p=0.009 and a statistically significant interaction between *distractor* and *distractor gender*: F(1, 405460)=8.81, p=0.003.

Tukey tests revealed reading times at the pronoun region were 189ms longer in sentences with distractors with definitional gender than in sentences with stereotypical gender ($p<0.005$).

The interaction plot illustrates the interaction reported above.

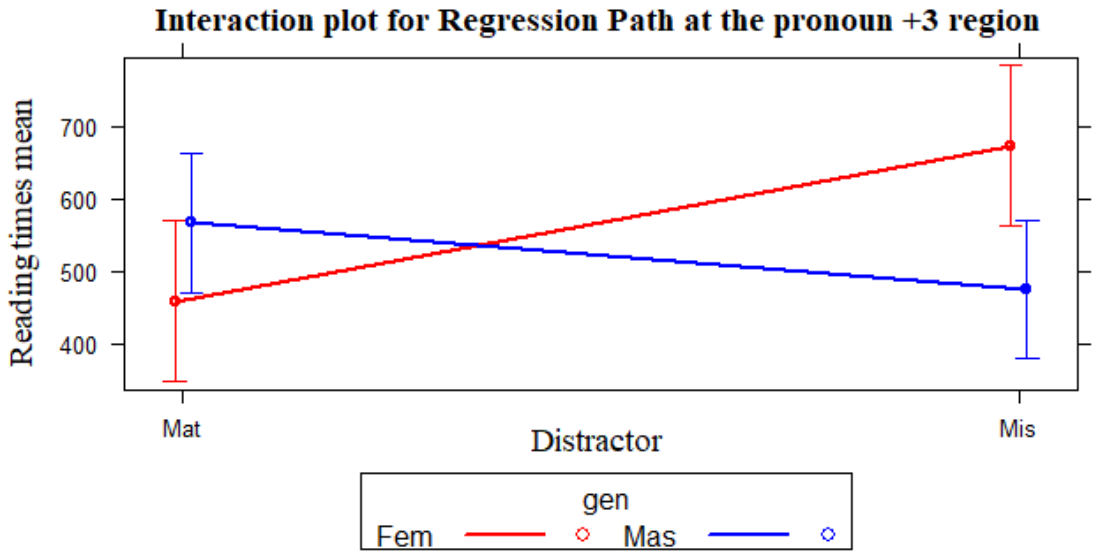

Figure 52: Interaction plot for Regression Path at the pronoun +3 region by distractor and gender in Experiments 2a and 2b

Tukey HSD tests detected reading times at the pronoun +3 region is 217ms faster for feminine matching distractors than for feminine mismatching distractors ($p=0.031$). Moreover, reading times were 201ms longer for mismatching masculine distractors than for mismatching feminine distractors.

Discussion

The results of Regression Path reading times suggest both structural and morphological cues are equally taken into account at the reanalysis/integration processing phases.

Antecedent retrievals were faster in sentences in which there were structurally acceptable antecedents that matched the pronoun in gender than in sentences in which there were mismatching structurally acceptable antecedents. This is evidence in favor of the importance of Principle B structural constraints. However, in sentences in which there was no matching structurally antecedent available, structurally unacceptable antecedents that matched the pronouns in gender would influence pronoun resolution, facilitating retrievals. This is evidence in favor of the fact similarity-based interference effects occur even in target mismatching conditions (Engelmann et al., 2015).

In addition, as expected, it seems definitional gender weighed more in memory than stereotypical gender. Definitional gender is preferably retrieved rather than stereotypical gender due to the fact the first might be lexically specified, while the latter would be a result of a more costly process in which probabilistic inferences are performed based on world-knowledge pragmatics.

In congruence with the predictions, masculine gender weighed more in memory than feminine gender. Masculine distractors caused inhibitory effects because they have increased prominence as it is the default gender. However, what is curious is that this happened for mismatching distractors, which means than even mismatching distractors can interfere in antecedent retrievals causing similarity-based interference effects.

Second Pass

Figure 53 corresponds to the Second Pass reading times at all regions of interest in Experiments 2a and 2b.

Tukey HSD tests showed reading times at the pronoun region were 108ms longer for Mis_Mat condition than for Mat_Mis condition ($p=0.028$); 156ms longer for Mis_Mis condition than for Mat_Mat condition ($p=0.001$); and 177ms longer for Mis_Mis than for Mat_Mis ($p<0.05$). Reading times at the pronoun +1 region were also 186ms for Mis_Mat condition than for Mat_Mis condition ($p=0.004$); 164ms longer for Mis_Mis condition than for Mat_Mat condition ($p=0.017$); and 238ms longer for Mis_Mis condition than for Mat_Mis condition ($p<0.05$).

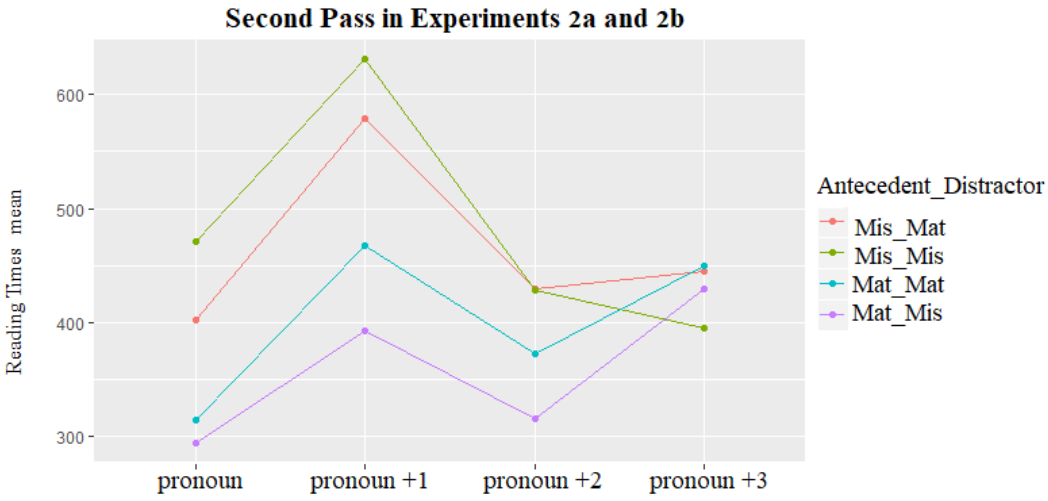

Figure 53: Line chart for Second Pass by regions of interest and conditions in Experiments 2a and 2b

ANOVA tests found statistically significant results for all regions of interest.

Pronoun region

The reading times and standard deviations at the pronoun region in milliseconds are reported in Table 40. Tukey HSD tests showed reading times were 194ms longer for Mis_Mis condition than for Mat_Mis condition for distractors with feminine

stereotypical gender (p=0.021); 191ms longer for Mat_Mis condition for distractors with feminine stereotypical than for Mis_Mis for distractors with feminine definitional gender (p=0.014)

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	422 (408)	385 (200)	447 (315)	382 (331)
Mis_Mis	483 (230)	456 (250)	461 (233)	484 (382)
Mat_Mat	378 (275)	287 (221)	215 (134)	345 (204)
Mat_Mis	416 (241)	229 (99)	307 (229)	255 (116)

Table 42: Second Pass reading times and standard deviations in parenthesis at the pronoun region in Experiments 2a and 2b

ANOVAs of the *lmes* for Second Pass at the pronoun region revealed a statistically significant main effect of *antecedent*: $F(1, 65505)=21.96$, $p<0.05$.

Tukey HSD tests indicated reading times at the pronoun region were 132ms longer for sentences with mismatching antecedents than for sentences with matching antecedents ($p<0.05$).

Pronoun +1 region

The reading times and standard deviations at the pronoun +1 region in milliseconds are reported in Table 41. Tukey HSD showed reading times at the pronoun +1 region were 258ms longer for Mis_Mat for distractors with feminine definitional gender than for Mat_Mat for distractors with feminine stereotypical gender (p=0.037) and 275ms for Mis_Mis than for Mat_Mis for distractors with feminine definitional gender (p=0.004).

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	522 (423)	690 (490)	639 (268)	461 (329)
Mis_Mis	756 (650)	648 (416)	494 (242)	599 (364)
Mat_Mat	483 (450)	575 (502)	318 (100)	386 (179)
Mat_Mis	538 (368)	360 (337)	345 (164)	357 (214)

Table 43: Second Pass reading times and standard deviations in parenthesis at the pronoun +1 region in Experiments 2a and 2b

ANOVAs of the *lmes* for Second Pass at the pronoun +1 region revealed a statistically significant main effect of *antecedent*: $F(1, 114605)=29.45$, $p<0.05$ and a statistically significant interaction between *distractor*, *distractor type of gender*, and *distractor gender*: $F(1, 114605)=5.32$, $p=0.021$.

Tukey HSD tests indicated reading times at the pronoun region were 177ms longer for sentences with mismatching antecedents than for sentences with matching antecedents ($p<0.05$). Moreover, reading times at the pronoun region were 124ms longer for sentences with definitional distractors than for sentences with stereotypical distractors ($p=0.001$).

The interaction plot in Figure 547 illustrates the interaction reported above for pronoun +1 region.

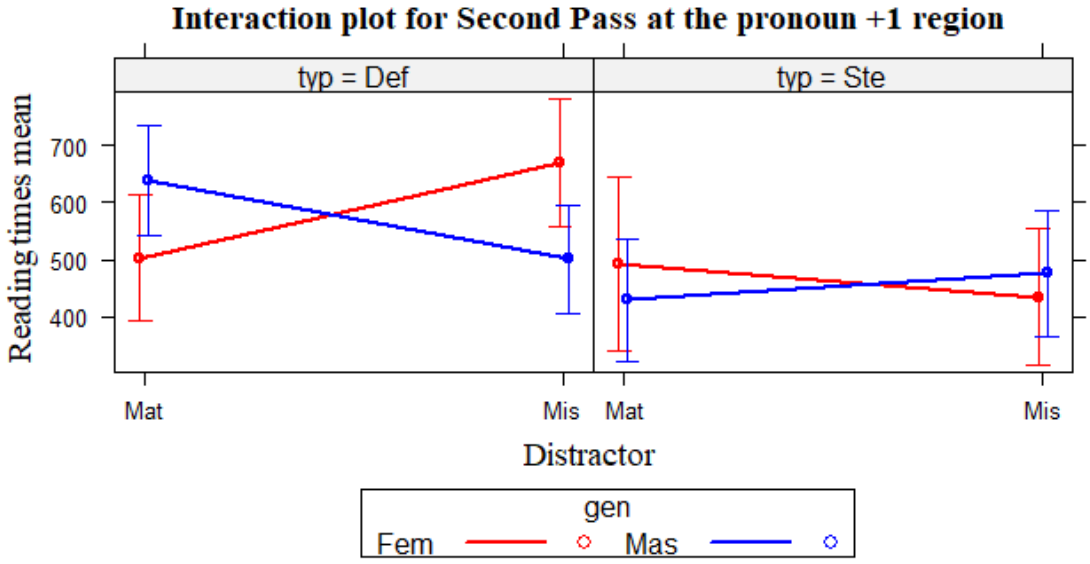

Figure 54: Interaction plot for Second Pass at the pronoun +1 region by distractor, type of gender, and gender in Experiments 2a and 2b

Tukey HSD tests detected reading times at the pronoun +1 region were 231ms longer for mismatching distractors with feminine definitional gender than for feminine stereotypical gender in a trend towards significance (p=0.088). Moreover, reading times at the pronoun +1 region were 210ms longer for matching distractors with masculine definitional gender than for matching distractors with masculine stereotypical gender in a trend towards significance (p=0.067).

Pronoun +2 region

The reading times and standard deviations at the pronoun +2 region in milliseconds are reported in Table 42. Tukey HSD tests did not show any statistically significant difference in Table 42.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	339 (236)	471 (294)	612 (480)	365 (149)
Mis_Mis	438 (252)	458 (304)	372 (220)	433 (404)

Mat_Mat	368 (205)	393 (209)	356 (214)	349 (232)
Mat_Mis	428 (429)	324 (233)	207 (115)	274 (136)

Table 44: Second Pass reading times and standard deviations in parenthesis at the pronoun +2 region in Experiments 2a and 2b

ANOVAs of the *lmes* for Second Pass at the pronoun +2 region revealed a statistically significant main effect of *antecedent*: F(1, 69192)=10.11, p=0.0016 and a statistically significant interaction between *distractor*, *distractor type of gender*, and *distractor gender*: F(1, 69192)=6.0, p=0.014.

Tukey HSD tests indicated reading times at the pronoun region were 84ms longer for sentences with mismatching antecedents than for sentences with matching antecedents (p=0.006).

The interaction plot in Figure 55 illustrates the interaction reported above for pronoun +2 region.

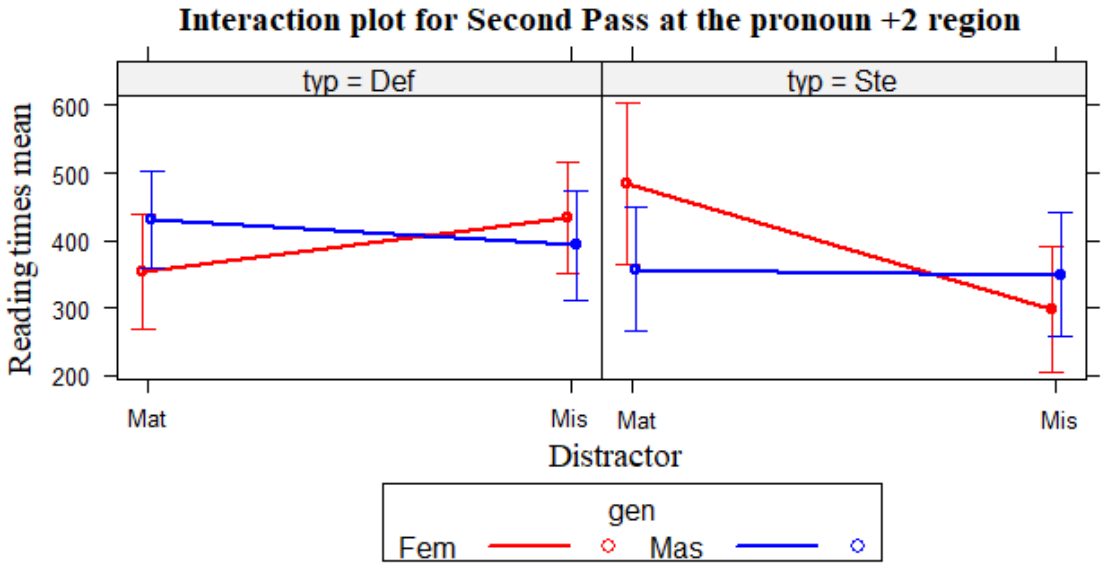

Figure 55: Interaction plot for Second Pass at the pronoun +2 region by distractor, type of gender, and gender in Experiments 2a and 2b

Tukey HSD tests detected reading times at the pronoun +2 region were 195ms longer for matching distractors with feminine stereotypical gender than for

mismatching distractors with feminine stereotypical gender in a trend towards significance (p=0.169).

Pronoun +3 region

The reading times and standard deviations at the pronoun +3 region in milliseconds are reported in Table 43. Tukey HSD tests did not show any statistically significant difference in Table 43.

	Definitional gender		Stereotypical gender	
	Feminine	**Masculine**	**Feminine**	**Masculine**
Mis_Mat	378 (214)	476 (395)	580 (428)	380 (275)
Mis_Mis	428 (296)	436 (299)	378 (297)	317 (216)
Mat_Mat	486 (378)	398 (237)	303 (176)	537 (442)
Mat_Mis	543 (560)	488 (408)	359 (263)	319 (206)

Table 45: Second Pass reading times and standard deviations in parenthesis at the pronoun +3 region in Experiments 2a and 2b

ANOVAs of the *lmes* for Second Pass at the pronoun +3 region revealed a statistically significant interaction between *antecedent*, *distractor type of gender*, and *distractor gender*: F(1, 90327)=5.70, p=0.017.

The interaction plot in Figure 56 illustrates the interaction reported above.

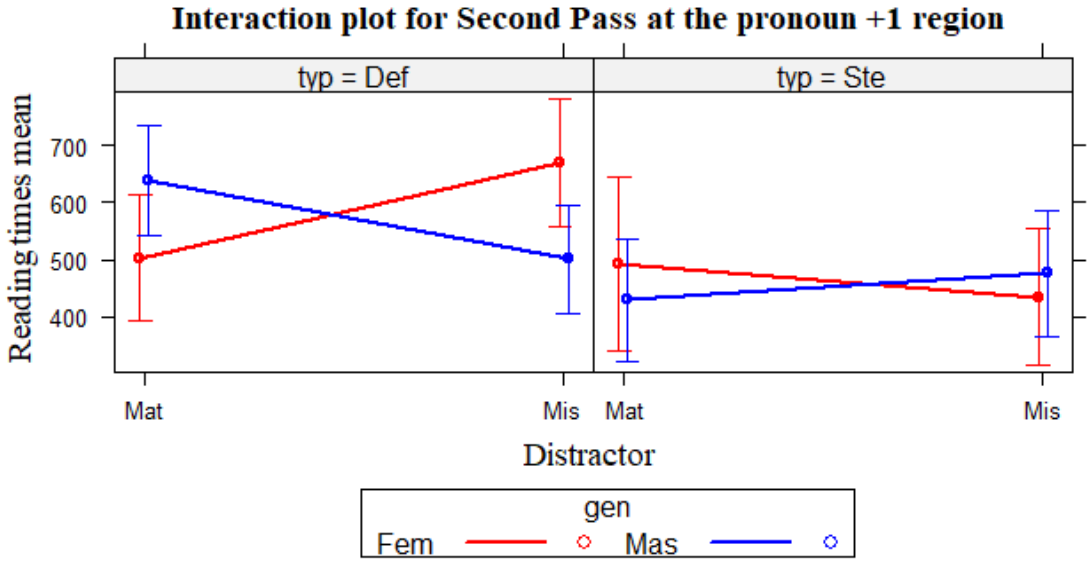

Figure 56: Interaction plot for Second Pass at the pronoun +3 region by antecedent, distractor, and gender in Experiments 2a and 2b

Tukey HSD tests did not detect any statistical significant difference in Figure 56.

Discussion

The results suggested both structural cues and gender morphological cues were taken into account in pronoun resolution at Second Pass, which corresponds to the latest processing stage. However, it seems structural cues seemed to play a major role than morphological cues.

As predicted, Principle B structural constraints cues were of great importance at antecedent retrievals once mismatching antecedents were responsible for more costly pronoun resolution than matching antecedents. As also predicted, definitional gender weighed more in memory retrieval than stereotypical gender. Distractors with definitional gender might be more prominent than distractors with stereotypical

gender as the former is lexically specified, while the latter is resulted from pragmatic world-knowledge inferences.

In addition, it appears matching distractors with feminine stereotypical gender causes slower antecedent retrievals than mismatching distractors. Thus matching distractors cause more similarity-based interference effects than mismatching distractors (Lewis & Vasishth, 2005).

4.4.3.6 Discussion of Experiments 2a and 2b

The results of Experiment 2a and 2b revealed pronominal structural constraints and gender morphological cues are both taken into account in antecedent retrieval in memory.

As predicted, antecedent retrievals were faster when structurally acceptable antecedents, which carry the [+ACCESSIBLE] feature (Kush & Phillips, 2015), agreed with the pronouns in gender than when they disagreed. This is evidence in favor of the idea not only structural binding constraints are important, but also agreement cues. However, structurally unacceptable antecedents were also taken into account in memory retrieval. For example, when the structurally unacceptable antecedent mismatched the pronouns in gender, the presence of structurally unacceptable antecedents that matched the pronouns in gender facilitated pronoun resolution. This means the distractors were misretrieved by memory as if they were the antecedents. This facilitation is called intrusion effects and it was already found out by Sturt (2003) and Chow et al. (2014), but not at early processing measures as it was found out here.

On the other hand, matching distractors can also be responsible for slower antecedent retrievals at late processing measures. In this case, matching distractors would cause similarity-based interference effects, competing with the antecedents in memory, causing slower retrievals. Similarity-based interference effects caused by matching distractors are predicted by the content addressable memory model (Lewis & Vasishth, 2005); however, it seems that even mismatching distractors can cause similarity-based interference effects.

As predicted, definitional gender weighed more in memory than stereotypical gender because the former is lexically retrieved, while the latter is pragmatically inferred through world knowledge probabilities (Carreiras et al., 1996; Oakhill et al., 2008; Kneiner et al., 2008; Canal et al., 2015). This way, distractors with definitional gender might be more prominent than distractors with stereotypical gender (Engelmann et al., 2015).

In congruence with the predictions, masculine distractors cause more similarity-based interference effects than feminine distractors. Masculine gender seems to the more prominent than feminine because it is the default gender in the language (Corbett, 1991; Côrrea et al., 2004; Casado et al., 2017; Lawall et al., 2012; Alves, 2014). Thus masculine distractors they cause both facilitatory and inhibitory effects. In the first case, intrusion effects caused by masculine distractors mean they were misretrieved by memory as if they were the antecedents; however, in the second case, they compete with the structurally acceptable antecedents, causing slower retrievals.

Similar to Experiments 1a and 1b, there were no major differences between the different eye-tracking measures and the different regions of interest in Experiments 2a and 2b. This corroborates to the idea that antecedent retrieval in

Brazilian Portuguese is a solid and constant process that simultaneously takes both structural and morphological cues into account throughout processing, with no exclusive early or late effects.

4.4.3.7 *Comparing the types of gender used in Experiments 1a/1b and in Experiments 2a/2b*

The materials used in Experiments 1a/1 were the same used in Experiments 2a/2b with the purpose of comparing the four types of gender tested: semantic gender, grammatical gender, definitional gender, and stereotypical gender.

The results of Tukey HSD of Second Reading Pass at the pronoun region detected reading times were: 101ms longer for distractors with definitional gender than for distractors with semantic gender ($p<0.05$); 91ms longer for distractors with stereotypical gender than for distractors with semantic gender ($p=0.004$); 98ms longer for distractors with stereotypical gender than for distractors with grammatical gender ($p=0.002$).

Thus considering the results found in Experiments 1a/1b and in Experiments 2a/2b together with the results presented above, it seems the four types of gender tested in the present work weigh in memory in the following order:

(90) grammatical gender < semantic gender < stereotypical gender < definitional gender.

Definitional gender might weigh the most (van Dyke & McElree, 2011) in memory because its gender information is lexically specified, which makes distractors with this type of gender very prominent. Following definitional gender, the second

type of gender than seem to weigh the most is stereotypical gender. Despite being probabilistically resulted from world-knowledge inferences, stereotypical gender seems to be intrinsically connected to the lexical information of the nouns. Although it is not as specified as definitional gender, stereotypical information might be calculated before the word is selected from the lexicon, which makes distractors with this type of gender very prominent as well. The third that weigh the most in the rank is semantic gender. Unlike definitional and stereotypical gender, compositional/derivational semantic gender is not lexically determined; it is retrieved by mental grammar in Stripping Affixes process. This type of gender requires additional processes in order to be retrieved, which makes them not so prominent. Despite the fact semantic compositional/derivational gender is more costly retrieved, it is more preferably than grammatical gender once semantic gender is conceptually motivated.

4.4.4 General Discussion

The main aim of the present work was to investigate the role of structural cues and gender morphological cues in pronominal antecedent retrieval in Brazilian Portuguese. Our results showed effects of an integration of both structural cues and gender cues in all of the eye-tracking measures investigated, which is evidence in favor of the idea that those cues work simultaneously together throughout the process of pronominal antecedent retrieval in memory (Chow et al., 2014).

More importantly, since our results showed structurally unacceptable antecedent candidates influence coreference processing since the beginning, it seems structural cues do not work as an initial filter blocking interference from structurally

unacceptable antecedent candidates. This contradicts the view that structural constraints work as an initial filter in binding processing [Nicol & Swinney, 1989; Clifton et al., 1997; Sturt, 2003; Leitão et al., 2008; Oliveira et al., 2012; Dillon et al., 2013; Chow et al., 2014; Cunnings et al., 2015]. On the contrary, the results reported here showed evidence in favor of the initial fallibility of structural binding constraints [Badecker & Straub, 2002; Kennison, 2003; Cunnings & Felser, 2013; Parker, 2014; Patil et al., 2016].

Furthermore, it should be highlighted that as the structural cues do not block the influence of structurally unacceptable antecedent candidates in coreference processing; similarly, the morphological cues do not block the influence of antecedent candidates that mismatch those morphological cues. The results reported here showed that even morphologically mismatching antecedent candidates could influence antecedent retrievals.

Thus, since neither the structural cues nor the morphological cues work as an initial filter blocking interference from, for example, structurally unacceptable candidates or gender mismatching candidates, the results reported in this dissertation seem to refute the predictions of the content-addressable memory model (Lewis & Vasishth, 2005). According to this model, only the items that content match the target can be taken into account by memory; therefore, mismatching distractors, which neither match the structural cues nor the morphological cues would be disregarded by memory. However, the results of this dissertation showed this is not the case, that is, both matching and mismatching distractors can interfere in memory retrieval process. It appears memory takes every single antecedent candidate into account: those matching or mismatching the structural cues and those matching or mismatching the morphological cues. This strategy is not so demanding as it sounds since activation

occurs in parallel, which means the number of candidates does not affect the retrievals speed (McElree, 2000). In addition, it prevents costly reanalysis processes in case no antecedent is found in the given context, which happens more often than one might think. Although, the number of candidates activated does not affect the retrievals speed, it decreases the chances of the most adequate candidate to be retrieved. This is predicted by the content addressable memory model, which posits the number of candidates decrease the strength of activation among the items, causing misretrievals and cue confusion. Misretrievals generate facilitatory effects and cue confusion generates inhibitory effects (see section 2.6.4).

The results of the present dissertation showed distractors can cause either facilitatory or inhibitory effects. In the first case, retrievals are faster because distractors might be misretrieved in memory as if they were the antecedents; while in the second case, retrievals are slower because distractors might compete with the antecedents in memory. According to Engelmann et al. (2015), in both facilitatory and inhibitory situations, candidates need to be prominent in discourse.

Since all distractors were the subjects of a relative clause, it seems the materials used in the present work maximized the discursive prominence of structurally unacceptable antecedents (Gordon & Hendrick, 1998; Grosz et al., 1995; Arnold et al., 2000; Foraker & McElree, 2007; Rigalleau et al., 2004). However, it seems that more than discursive prominence, what explains our results is cue prominence in memory, which brings the discussion to the experimental hypotheses of this dissertation.

The first hypothesis was gender morphological cues would be of paramount importance in pronominal antecedent retrieval in Brazilian Portuguese since it is a language in which readers might be used to rely on surface cues in order to process

language in real time (Lago, 2014). Thus it was expected that even structurally unacceptable candidates violating Principle B would be considered as potential antecedent candidates if they agreed in gender with the pronouns, despite the fact Principle B would be violated, as these candidates would be located in the same governing category of the pronouns. The results presented here confirmed the first hypothesis of this dissertation, that is, morphological cues were so crucial for pronominal antecedent retrievals that memory seemed to be influenced by all gender cues displayed in the discursive context, even the ones displayed by the candidates that mismatched the pronouns. Nevertheless, binding structural constraints were also taken into account in antecedent retrievals. Structurally acceptable candidates, carrying the [+ACCESSIBLE] c-command feature, that agreed with the pronouns in gender were responsible for facilitating antecedent retrievals when compared to those that disagreed with the pronoun in gender.

The second hypothesis was that memory would be influenced by different types of gender cues. Thus antecedent candidates would be encoded/retrieved at different weights depending on the type of gender they convey. The results reported also confirmed the second hypothesis of this dissertation, that is, different types of gender have different weight is memory retrieval. More prominent types of gender would weigh more in memory, and consequently, they would be more preferably retrieved.

The results found in Experiments 1a and 1b showed compositional/derivational semantic gender is more prominent in memory than grammatical gender; as a result, distractors with semantic gender would cause more similarity-based interference effects, that is, they would distract memory more than distractors with grammatical gender, causing slower retrievals. Semantic gender

would be more preferably retrieved than grammatical gender because the former is conceptually motivated (Vigliocco & Franck, 1999). The results found in Experiment 2a and 2b revealed definitional gender is more prominent in memory than stereotypical gender, since distractors carrying definitional gender were responsible for slower retrievals than distractors carrying stereotypical gender. Definitional gender weighed more in memory than stereotypical gender because the former is lexically specified, while the latter is inferred from world-knowledge probabilities (Carreiras et al., 1996; Oakhill et al., 2008; Kneiner et al., 2008; Canal et al., 2015). When Experiments 1a and 1b are compared to Experiments 2a and 2b, the results pointed out the slowest retrievals occurred for distractors with definitional gender, followed by distractors with stereotypical gender, followed by distractors with compositional/derivational semantic gender, followed by distractors with grammatical gender. This order indicates the order of prominence in memory of each type of gender investigated.

The Pre-tests showed grammatical gender might be conceptually biased towards masculine or feminine; on the contrary, Experiments 1a and 1b revealed grammatical gender seemed not to be conceptually biased, as they were processed distinctly from semantic gender, which is a truly conceptually motivated gender. A reason for this discrepancy between the Pre-tests and the Experiments 1a and 1b might be related to the fact that grammatical gender only activate their conceptual bias when required by the task, and since Experiments 1a and 1b did not require any comprehension question about the referents' sex, this bias was not activated.

Since Pre-Test 1 did not show any on-line differences between stereotypical and grammatical gender, it seems that stereotypical gender might be specified in the lexicon just like grammatical gender is. However, the results of Pre-Test 2 showed it

is easier to find a conceptual referent for nouns with stereotypical gender than for nouns with grammatical gender. These results suggests that although, both grammatical gender and stereotypical gender seem to be lexically specified, the former seems to be underspecified for gender, while the former seems to be gender-determined. With respect to the eye-tracking experiments, the results pointed out stereotypical gender weighed less in memory than definitional gender, since distractors with definitional gender were responsible for slower retrievals than distractors with stereotypical gender. Thus, it appears stereotypical gender also has similitudes and differences with definitional gender, which is another type of gender that is specified in the lexicon. Their gender information can be both determined in the lexicon; however, while definitional gender seems to be gender-determined, stereotypical gender seems to be underspecified for gender requiring pragmatic inferences based on world-knowledge probabilities. This way, definitional and stereotypical genders have different weights in memory.

In congruence with the Pre-tests, Experiments 1a/1b and Experiments 2a/2b detected masculine gender is more prominent than feminine gender, that is, masculine gender weighs more in memory than feminine gender. An evidence for that lies in the fact that masculine distractors cause more similarity-based interference effects than feminine distractors. Masculine distractors were responsible for slower retrievals than feminine distractors, because masculine gender is the default gender in Portuguese (Corbett, 1991; Côrrea et al., 2004; Casado et al., 2017; Lawall et al., 2012; Alves, 2014), that is, it is more preferably to be retrieved than feminine.

It should be mentioned distractors with more prominent cues in memory might cause either inhibitory or facilitatory effects. The presence of prominent genders such as semantic gender and masculine gender can either speed up or slows down

coreference processing. It depends on the features of the other antecedent candidates involved. Therefore, the next step of the present work is to build a model capable of predicting whether distractors would cause inhibitory or facilitatory effects in antecedent retrievals.

The present work provided novel evidences on the influence of gender cues in pronominal antecedent retrieval. The results reported here corroborate to the idea that memory takes into account all information available in the context, including those conveyed by structurally unacceptable candidates that gender mismatch the pronouns. And since the content addressable memory model states only candidates that match or partial match the distractors are considered as potential items to be retrieved, this model could not explain why mismatching distractors could interfere in antecedent retrievals. Moreover, this model does not predict features might have different weights in memory, which might mean some candidates may be more prominent than others because of the features they carry, including gender features. Thus the next step is to present amendments to the content addressable memory model in order to accommodate the results reported here.

5. Summary and conclusions

This dissertation aimed at understanding how pronouns retrieve their antecedents in memory, and more specifically, how gender cues affect this process in a language with redundant visible morphological marks such as Brazilian Portuguese. Given the fact that readers/speakers of Brazilian Portuguese are used to rely on morphological cues in order to process language in real time, it was hypothesized that gender morphological cues would play a great role in antecedent retrieval. Thus it was predicted to find effects of gender morphological cues throughout coreference processing, that is, from initial to late processing stages. In addition, it was predicted that structurally unacceptable antecedent candidates that agreed in gender with the pronouns would be considered as potential antecedents by memory, despite the fact they violate binding structural constraints.

In order to understand the role of gender morphological cues in pronominal antecedent retrieval, it was necessary to dissociate gender cues from binding structural constraints. Thus, the first hypothesis was tested by manipulating the type of gender conveyed by the structurally unacceptable antecedent candidates. The distractors could agree in gender with the pronouns, despite the fact they violate Principle B structural constraints.

Since the type of gender conveyed by distractors was manipulated, it seemed to be reasonable to investigate which types of gender could interfere more in antecedent retrieval, that is, which types of gender would be more prominent in memory. This leads to the second hypothesis of this dissertation, which claims memory is able to distinguish between different types of gender, so that antecedent candidates would be retrieved by memory according to the type of gender they would

convey. In other words, different types of gender would be encoded/retrieved in memory with different weights so that candidates with more prominent gender features would be preferably retrieved. It was expected semantic gender would weigh more in memory than grammatical gender since the former is both conceptually and syntactic motivated, while the latter is only syntactic motivated. Furthermore, it was predicted definitional gender would weigh more than stereotypical gender because the former is lexically determined, while the latter is a result of probabilistic inferences based on world-knowledge. It was also expected masculine gender would weigh more than feminine gender due to the fact masculine is the default gender, that is, it is a generic gender that holds both masculine and feminine representations.

In order to test the hypotheses, two pre-tests and four eye-tracking experiments were conducted with native speakers of Brazilian Portuguese. Below, one can find a summary of the main findings of this dissertation:

(91) Pre-test 1: agreement cloze task

(a) On-line reaction times indicated grammatical gender and stereotypical gender were equally gender assigned; consequently, it seems stereotypical gender is lexically specified in the noun's representation as grammatical gender is;

(b) Neutral bigenders were more gender assigned as masculine than feminine, since masculine is the default gender. For example, *banhista* (bather), *hóspede* (guest), *informante* (informant), *repórter* (reporter), *romancista* (romanticist), *sem-teto* (homeless person);

(c) Feminine epicenes were more gender assigned as feminine than masculine epicenes as masculine, feminine epicenes are more marked than masculine epicenes;

(d) Some epicenes were treated like bigenders, receiving ambiguous (a/o) gender assignment. For example, *bebê* (baby), *cônjuge* (spouse), *dedo duro* (snitch), *neném* (baby);

(e) Male participants gender assigned masculine epicenes more ambiguously than female participants, that is, female participants seemed more conservative.

(92) Pre-test 2: gender bias judgment

(a) Masculine stereotypical nouns were judged faster than masculine grammatical gender, which might mean stereotypical nouns are underspecified for gender, requiring pragmatic inferences based on world-knowledge probabilities in order to retrieve their gender information, while epicenes would be conceptually specified ambiguously;

(b) Feminine epicenes were judged faster than masculine epicenes, which might mean masculine epicenes are less marked than feminine epicenes;

(c) Masculine stereotypical gender were judged faster than feminine stereotypical gender, since masculine is the default gender;

(d) Masculine nouns were judged slower than feminine nouns as masculine might carry both masculine and feminine representations;

(e) Epicenes were judged as more masculine than bigenders, which might mean default gender affects more epicenes;

(f) Male participants were more masculine biased than female participants.

(93) Experiments 1a and 1b: compositional/derivational semantic gender *versus* grammatical gender

(a) Both structural cues and gender cues play a role throughout coreference processing as soon as the pronoun is encountered;

(b) Structural cues does not work as an initial filter blocking the interference of structurally unacceptable candidates, even when there is a structurally acceptable antecedent available in the sentence that matches the pronoun;

(c) Semantic gender weighed more in memory than grammatical gender, because semantic gender is redundant, that is, it carries both syntactic and conceptual gender, while grammatical gender only carries syntactic gender;

(d) Masculine gender weighed more in memory than feminine gender, because masculine gender is the default gender;

(e) Structurally unacceptable candidates with prominent genders such as semantic gender and masculine gender can cause either facilitatory (in this case, they are misretrieved) or inhibitory effects (in this case, they compete with the structurally acceptable antecedents).

(94) Experiments 2a and 2b: definitional gender *versus* stereotypical gender

(a) Both structural cues and gender cues play a role throughout coreference processing as soon as the pronoun is encountered;

(b) When the structurally acceptable antecedents mismatched the pronouns in gender, structurally unacceptable antecedents that matched the pronouns in gender facilitated coreference processing because they were probably misretrieved;

(c) Structurally acceptable antecedents that either match or mismatch the pronouns can also cause inhibitory effects, competing with antecedents;

(d) Definitional gender weighed more in memory than stereotypical gender, because the former is lexically determined, while the latter requires pragmatic inferences based on world-knowledge probabilities;

(e) Masculine gender weighed more in memory than feminine gender, because masculine gender is the default gender.

(95) Experiments 1a/1b *versus* Experiments 2a/2b

(a) Prominence order in memory: grammatical gender < semantic gender < stereotypical gender < definitional gender

Based on the results of the pre-tests and the experiments, this dissertation corroborated in favor of the first experimental hypothesis of this dissertation. In other words, gender morphological cues are so important for antecedent retrieval in Brazilian Portuguese that they play a role in each and every step of coreference processing, from early to late stages. Not only gender morphological cues are crucial for antecedent retrieval, but also the structural cues. This dissertation showed both structural constraints and gender morphological cues are equally and simultaneously involved in pronominal antecedent retrieval in memory (Chow et al., 2014). However, they do not work as filters, that is, memory takes into account both structurally unacceptable candidate antecedents and gender mismatching candidates. These results contradict the content-addressable memory model (Lewis & Vasishth, 2005), which states that only candidates that match [or partial match] the target contents are activated in memory.

Moreover, the results showed in this dissertation also corroborates in favor of the second experimental hypothesis, that is, memory can indeed be influenced by

language. In other words, different gender cues seem to have different weights (van Dyke & McElree, 2011) in memory so that antecedent candidates with more prominent gender are preferably retrieved in memory. The most prominent gender in memory seems to be definitional gender (lexically determined), followed by stereotypical gender (although seems to be lexical, it depends on world-knowledge inferences), compositional/derivational semantic gender (conceptually motivated), and grammatical gender (only syntactically motivated). It should be mentioned the content addressable memory model couldn't explain different weights in memory that, for example, gender cues might have. This way, this model needs to be adjusted in order to contemplate the results presented here.

Finally, this dissertation could answer some of the questions raised on page 150:

(a) nouns with grammatical gender that refer to humans seem to be semantically motivated (Vigliocco et al., 2005)? To which extent are they different from nouns with semantic gender?

Nouns with grammatical gender that refer to humans seem to be conceptually biased depending on the task involved. In Pre-Test 2, which was a rating scale study aiming at investigating the conceptual referents of the nouns tested, it was noticed masculine epicenes were masculine biased, while feminine epicenes were feminine biased. However, in the eye-tracking experiments, there was no evidence in favor of the conceptual bias involving grammatical gender. Nouns with grammatical gender and nouns with semantic gender were treated differently by memory, that is, nouns with semantic gender seem to be more preferably to be retrieved than nouns with grammatical gender. This means that semantic gender weighs more in memory than

grammatical gender, that is, semantic gender is more prominent in memory than grammatical gender.

(b) is compositional semantic gender more processing demanding than semantic definitional gender? (Allen et al., 2003)

Although the experiments conducted by this dissertation did not test this particular question, it was possible to learn definitional gender might weigh more in memory than compositional gender. Definitional gender seems to be more prominent in memory due to the fact its gender information is lexically determined, and not morphologically determined as it happens for compositional gendered nouns. This way, structurally unacceptable candidates with definitional gender would be responsible for more interference effects than those structurally unacceptable candidates with compositional gender.

(c) is stereotypical gender similar to definitional gender? (Kreiner et al., 2008; Osterhout et al., 1997; Canal et al., 2015)

It seems both stereotypical and definitional genders are lexically determined; however, through distinct processes. The former might be specified by probabilistic inferences based on world knowledge, while the latter might be specified by a language idiosyncrasy. Despite their similarity, they work in memory distinctly, that is, definitional gender might weigh more in memory than stereotypical gender. A reason for definitional gender being more prominent in memory than stereotypical gender is probably related to the fact it is an intrinsic linguistic property and not a result from world knowledge inferences.

(d) does redundancy of surface cues facilitate processing? (Cacciari et al., 1997)

Since the experiments conducted in the present dissertation did not test this particular question, it is difficult to answer it. However, our results indicate structurally unacceptable antecedents with redundant surface cues weigh less in memory, causing less interference in memory.

(e) is masculine the default gender in two-gender Romance languages, that is, does masculine gender evoke both masculine and feminine representations? (Casado et al., 2017);

The results found in both Pre-tests and eye-tracking experiments revealed masculine gender seems to be the default gender in Portuguese, which means it might evoke both masculine and feminine conceptual representations. Masculine gender can either speed up or slow down processing. In Pre-Test 1, masculine nouns were processed slower than nouns with feminine nouns. In the eye-tracking experiments, masculine gender seemed to weigh more in memory than feminine gender. Masculine gender might be more prominent in memory due to the fact it can carry both masculine and feminine conceptual representations.

(f) does the sex of the comprehenders always affects gender retrieval (Osterhout et al., 1997; Kennison & Trofe, 2003; Casado et al., 2017)?

Unfortunately, it was not possible to control for the sex of the participant in the eye-tracking experiments; however, the Pre-test 2 results suggested male participants seem to be more masculine biased, while female participants seem to be more feminine biased.

To conclude, this dissertation provided novel evidences in the field, showing that in languages with overt morphology, both discursive, structural and gender cues play a role in coreference processing; but differently from what was previously presented in the literature, neither structural cues nor morphological cues work as [initial] filters. It seems memory considers all possible candidates at the same time, so that candidates whose features weigh more in memory would be preferably retrieved. The results presented here showed that some of those prominent features in memory are [+ACCESSIBLE], which is related to structural constraints; [+ LEXICAL], which is related to definitional gender; [+ SEMANTIC], which is related to conceptual gender; and [+TOPIC], which is related to discourse salience.

The take home message of this dissertation is a language with overt morphology such as Brazilian Portuguese can shed light on the processes that underlie antecedent retrievals in memory. The results reported here showed [gender] morphological cues have paramount importance in coreference processing, and unlike previous works in the literature, there were found evidences that memory is capable of differentiating from different types of gender cues in order to quickly retrieve the antecedents. This sophisticated mechanism directly dialogues with the linguistic nuances that make some cues to be more preferably to be retrieved than others. More preferably cues weigh more in memory, that is, they are more prominent in memory, and consequently, they can be more easily retrieved. This way, it seems language and memory are more closely connected than we thought.

References

Afonso, O., Domínguez, A., Álvarez, C., Morales, D. (2013). Sublexical and lexico-syntactic factors in gender access in Spanish. *J. Psycholinguist Res*.

Alcocer, P., Phillips, C. (2012). *Using relational syntactic constraints in content-addressable memory architectures for sentence parsing*. University of Maryland.

Allen, M., Badecker, W., Osterhout, Lee. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes*, 18 (4), 405-430.

Alves, M. (2014). *Processamento do traço de gênero na correferência pronominal com antecedentes sobrecomuns e comuns de dois gêneros no português do Brasil*. Dissertação de mestrado. Rio de Janeiro: Universidade Federal do Rio de Janeiro.

Anderson, J. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behaviour*, 22.

Anderson, J. R. & Lebiere, C. (1998). *Atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Anderson, J. R., Bothell, d., Byrne MD. Douglass, S., Lebiere, C. & Qin, Y. (2004). An integrated theory of the mind. *Psychology Review*, Oct; 111 (4), 1036-1060.

Anderson, S. (1982). Where's morphology? *Linguistic Inquiry*, vol. 13, no. 4, 571-612.

Arnold, J., Eisenband, J., Brown-Schmidt, S., Trueswell, J. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76.

Atkinson, R., Shiffrin, R. (1968). Chapter: Human memory: A proposed system and its control processes. In: Spence, K. & Spence, J. *The psychology of learning motivation*, vol.2. New Academic Press, 89-195.

Baddeley, A., Hitch, G. (1974). Working memory. In: G.H. Bower (Ed.). *The psychology of learning and motivation: Advances in research and theory*, vol. 8, p.47-89.

Badecker,W. & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphora. *J. Exp.Psychol.Learn.Mem. Cogn.* 28, p.748–769.

Bates, E., Devescovi, A., Pizzamiglio, L. D'Amico, S., Hernandez, A. (1995). Gender lexical access in Italian. *Perception & Psychopysics*, 57 (6), 847-862.

Bruening, B. (2006). *What is the Right Binding Theory?*. SUNY Stony Brook.

Butterworth, B. (1983). Lexical representation. In: B. Butterworth (Ed.), *Language Production*, London: Academic Press, 257-294.

Cacciari, C., Carreiras, M. & Cionini, C. B. (1997). When words have two genders: anaphor resolution for Italian fuctionally ambiguous words. *Journal of Memory and Language*, 37, 517-532.

Cacciari, C., Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns. *Applied Psycholinguistics,* 28, 277-293.

Cacciari, C., Corradini, P.; Padovani, R. & Carreiras, M. (2011). *Pronoun resolution in Italian: the role of grammatical gender and context*. In: *Journal of Cognitive Psychology*. London, UK: Psychology Press.

Caffarra, S., Janssen, N., Barber, H. (2014). Two sides of gender: ERP evidence for the presence of two routes during gender agreement processing. Neuropsychologia, 63, 124-134.

Caffarra, S., Siyanova, A. , Pesciarelli, F., Vespignani, F., Cacciari, C. (2015). Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian. *Psychophysiology*, 52, 1019-1030.

Canal, P., Garnham, A., Oakhill, J. (2015). Beyond gender stereotypes in language comprehension: self sex-role descriptions affect the brain's potentials associated with agreement processing. *Frontiers in Psychology* 6:1953.

Caramazza, A., Laudanha, A., Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28, 297-332.

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*, 49A (3), 639-663.

Casado, A., Palma, A., Paolieri, D. (2017). The influence of sex information on gender word processing. *Journal of Psycholinguistic Research*, published online.

Chomsky, N. (1970). *Remarks on nominalization*. In: Jacob, R., Rosenbaum, P. (Eds.) Readings in English transformational grammar, Waltham, Massachusetts: Blaidsdell.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

_____. (1986). *Knowledge of Language*. Praeger, Conneticut.

Chomsky, N. (1993). A minimalist program for linguistic theory. In: Hale, K. Keyser, S. J. The view from building 20. Cambridge, Massachusetts: The MIT Press, 1-52.

Chomsky, N. & Lasnik, H. (1993). *The Theory of Principles and Parameters*. In: Jacobs, J et alii. (org). *Syntax, an International Handbook of Contemporary Research*. Walter de Gryter, Berlin, New York.

Chow, W.; Shevaun, L. and Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*. 5, 2014, article 630.

Clifton, C., Frazier, L., & Deevy, P. (1997). Feature manipulation in sentence comprehension. *Rivista di Linguistica*, 11, p.11–39.

Corbett, G. (1991). *Gender*. Great Britain: Cambridge University Press.

Corrêa, L. (1998). *Acessibilidade e paralelismo na interpretação do pronome sujeito e o contraste pro/pronome em português*. DELTA: São Paulo.

Corrêa, L. (2001). Concordância de gênero no processamento de formas pronominais. *Cadernos de Estudos Linguísticos*, Campinas, 40, 77-92.

Corrêa, M., Almeida, D., Porto, R. (2004). On the representation of Portuguese gender-inflected words in the mental lexicon. *Brain and Language*, 90, 63-73.

Cowan, N. (2001). The magical number four in short-term memory: A consideration of mental storage capacity. Behavioral and Brain Sciences, 24, 87-185.

Cunnings, I. & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, 28:1-2, 188-219.

Cunnings, I., Patterson, C. & Felser, C. (2015). Structural constraints on pronoun binding and coreference: evidence from eye movements during reading. *Frontiers in Psychology*, 6:840.

Di Sciullo, A., Williams, E. (1987). *On the definition of wo*rd. Cambridge, Massachusetts: The MIT Press.

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, p.85–103.

Duff, S., Keir, J. (2004). Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge. *Memory and Cognition*, 32 (4), 551-559.

Dupoux, E. (1998). Words and how we (eventually) find them. In: ALTMANN, G. T.M. *The ascent of Babel: an exploration of language, mind, and understanding*. Oxford University Press, 65-83.

Engelmann, F., Jäger, L., Vasishth, S. (2015). The determinants of retrieval interference in dependency resolution: Review and computational modeling. *Journal of Memory and Language*, submitted.

Foraker, S., McElree, B. (2007). The role od prominence in pronoun resolution: active versus passive representations. *Journal of Memory and Language*, 56, 357-383.

Gabrieli, J. D. E., Brewer J. B., Desmond, J. E., Glover, G.H. (1997). Separate neural bases of two fundamental memory processes in the human medial temporal lobe. Science, 276, 264-266.

Garcia, D. (2009). *Elementos estruturais no acesso lexical: o reconhecimento de palavras multimorfêmicas no português brasileiro*. Master thesis, UFRJ.

Gordon, P., Grosz, B., Gilliom, L. (1993). Pronouns, names, and the centering of attention in discourse. Cognitive Science, 17, 311-347.

Gordon, P., Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science*, 22, 389-424.

Gordon, P., Hendrick, R., Ledoux, K, Yang, C. (1999). Processing of reference and the structure of language: An analysis of complex noun phrases. *Language and Cognitive Processes, 14, 353-379.*

Greene, S. B., McKoon, g. & Ratclif, R. (1992), Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 266-283.

Grosz, B., Weinstein, S., Joshi, A. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, vol. 21, n. 2.

Halle, M., Marantz, A. (1993). Distributed Morphology and the pieces of inflection. In: Hale, K. Keyser, S. J. The view from building 20. Cambridge, Massachusetts: The MIT Press, 111-176.

Halle, M., Marantz, A. (1994). Some keys of Distributed Morphology. MITWPL 21: Papers on phonology and morphology, 275-288.

Hobbs, J. (1979). Coherence and Coreference. *Cognitive Science*, 3, 67-90.

Hora, K. (2014). *O processamento da correferência pronominal anafórica em estruturas complexas do português brasileiro*. Dissertação de mestrado. Rio de Janeiro: Universidade Federal do Rio de Janeiro.

Kehler, A. (2007). Rethinking the SMASH approach to pronoun interpretation. In J. Gundel & N. Hedberg (Eds.), *Interdisciplinary perspectives on reference processing*, 95-122. New Directions in Cognitive Science. Oxford University Press.

Kehler, A., Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. Theoretical Linguistics, 39, 1-37.

Kenedy, E. (2016). Interfaces. In: *Curso Básico de Linguística Gerativa*, 1a ed. São Paulo: Contexto.

Kennison, S. M. (2003). Comprehending the pronouns *her*, *him*, and *his*: Implications for theories of referential processing. *Journal of Memory and Language*, 49, 335–352.

Kennison, S., Trofe, J. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, vol. 32, no. 3.

Kneiner, H., Sturt, P., Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: evidence from eye-movements. *Journal of Memory and Language*, 58, 239-261.

Konishi, T. (1993) The semantics of grammatical gender: a cross-cultural study. *Journal of Psycholinguistic Research*, vol. 22, no. 5.

Kuno, S. (1987). *Functional syntax: anaphora discourse and empathy*. Chicago, IL: University of Chicago Press.

Kush, D., Lidz, J., Phillips, C. (2015). Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language*, 82, 18-40.

Lago, M. S. (2014). *Memory and prediction in cross-linguistic sentence comprehension*. PhD dissertation, University of Maryland.

Lawall, R., Maia, M. & Amaral, L. (2012). *Resolução pronominal com antecedentes sobrecomuns e comuns de dois gêneros em Português Brasileiro como lingual maternal, lingual de herança e como segunda lingual*. In: *Revista Linguística* – Revista do Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro. Rio de Janeiro: 2005 v. 8, n.2.

Leitão, M. (2005). *O Processamento do objeto direto anafórico no Português Brasileiro*. 2005. 160f. Tese de Doutorado. Rio de Janeiro: Universidade Federal do Rio de Janeiro.

_____. (2010). *Animacidade e paralelismo estrutural no processamento da correferência*. In: *Revista Linguística* – Revista do Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro. Rio de Janeiro: v. 6, n.1.

Leitão, M., Ribeiro, A., Maia, M. (2012). Penalidade do nome repetido e rastreamento ocular em português brasileiro. *Revista Linguística*, vol.8, n.2.

Leitão, M.; Peixoto, P & Santos, S. (2008) Processamento da correferencial intrasentencial em PB. In: *Veredas Online* 2, 50-61.

Lessa, P. (2014). O papel do ordenamento dos elementos referencialmente dependentes no processamento da correferência em português do Brasil. Dissertação de Mestrado. Rio de Janeiro: Universidade Federal do Rio de Janeiro.

Levshina, N. (2015). *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins Publishing House.

Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language* 59, 200-222.

Lewis, R. L. & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science* 29, 375–419.

Lewis, R.L., Vasishth, S., and Van Dyke, J.A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.(Regul.Ed)10,* 44–54.

Machado, A. L. (2016). *Resolução de ambiguidade no âmbito da correferência pronominal: um estudo psicolinguístico sobre a influência do contexto*. Dissertação de mestrado. Rio de Janeiro: Universidade Federal do Rio de Janeiro.

Maia, M., Lemle, M., França, A. (2007). Efeito stroop e movimento ocular no processamento de palavras. *Ciências e Cognição*, vol. 12, 02-17.

Maia, M. (2013). *Linguística experimental: aferindo o curso temporal e a profundidade do processamento*. Rev. Est. Ling., Belo Horizonte, v. 21, n. 1, 9-42.

Maia, M.; Garcia, D. & Oliveira, C. (2012). *The processing of conceptual anaphors and fully specified pronouns in intra-sentential contexts in Brazilian Portuguese*. In: *Revista Virtual de Estudos da Linguagem*, v. 6, 200-219.

Maia, M., Ribeiro, A. (2015). Jabuticaba liboramima lê mais fácil do que jornaleiro norbalense: um estudo de rastreamento ocular de palavras e pseudopalavras mono e polimorfêmicas. In: Buchweitz, A., Motta, M. (Org.). *Linguagem e Cognição,* 1ed, Porto Alegre: EDIPUC-RS, 143-154.

Mattoso Câmara Jr., J. (1970). *Estrutura da Língua Portuguesa*. Petrópolis: Vozes.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25, 71-102.

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory and Cognition*, 24, 173-187.

_____. (2000). Sentence comprehension is mediated by content addressable memory structures. *Journal of Psycholinguistic Research*, 29, 111–123.

_____. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 817-835.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67–91.

Nichols, E. A., Kao, Y-C., Verfaellie, Gabrieli, J. D. E. Working memory and long-term memory for faces: Evidence from fMRI and global amnesia for involvement of the medial temporal lobes. *Hippocampus*, 16, 604-616.

Nicol, J. and Swinney, D.A. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguists. Res.* 18, 5–19.

Oakhill, J., Garnham, A., Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory and Cognition*, 33 (6), 972-983.

Oberauer, K. & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55, 601-626.

Oliveira, R.; Leitão, M. & Henrique, J. (2012). A influência dos antecedentes vinculados e não vinculados no processamento da anaphora "a si mesmo". In: *Revista Linguística*, v.8, n.2.

Osterhout, L., Bersick, M, McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory and Cognition*, 25 (3), 273-285.

Parker, D. (2014). *The cognitive basis for encoding and navigating linguistic structure*. PhD dissertation, University of Maryland.

Patil, U., Vasishth, S., Lewis, R. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers of Psychology*, 7-329.

Pearlmutter, D. (1988). The split-morphology hypothesis: evidence from Yiddish. In: Hammond, M., Noonan, M. (Ed). *Theoretical morphology: approaches in modern Linguistics*. Orlando: Academic Press.

Pinker, S. (1991). Rules of Language. *Science*, vol. 253, 530-535.

Phillips, C., Wagers, M. W., Lau, E. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. *Syntax & Semantics*, 37, 147-180.

Pollard, C., Sag, I. (1992). Anaphors in English and the scope of Binding Theory. *Linguistic Inquiry*, 23.2, 261-303.

Queiroz, K., Leitão, M. (2008). Processamento do sujeito anafórico em português brasileiro. *Veredas on-line*. Juiz de Fora: UFJF, vol.2.

Ranganath, C., Blumenfeld, R.S. (2005). Doubts about dissociations between short- and long-term memory. *Trends in Cognitive Science*, Aug.9 (8), 374-380.

Reinhart, T. & Reuland, E. (1993). "Reflexivity". In: *Linguistic Inquiry*, 24, 657-720.

Resende, N. (2015). *Processamento e representação do gênero gramatical do português: um mecanismo associativo*. Doctoral Dissertation, UFSC.

Reuland, E. (2006). Binding Theory: Terms and concepts. In: *Binding Theory*.

Rigalleau, F., Caplan, D, Baudiffier, V. (2004). New arguments in favor of an automatic gender pronominal process. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 57:5, 893-933.

Rumelhart, D., McClleland, J. (1985). On learning the past tenses of English verbs. *Parallel distributed processing: explorations in the microstructure of cognition: vol. 2. Psychological and biological models*. Cambridge, MA: Bradford Books/MIT Press.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, vol. 2 (9), 437-442.

Seidenberg, M., McClelland, J. (1989). A distributed, development model of word recognition and naming. *Psychological Review*, vol. 96, no.4, 523-568.

Shallice, T., Warrington, E. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly Journal of Experimental Psychology*, 22:2, 261-273.

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review* 99, 195-231.

Sturt, P. 2003 The time-course of the application of binding constraints in reference. *Journal of Memory and Language* 48: 542.

Taft, M., Forster, K. (1975). Lexical storage and retrieval of prefixed words. *Journal of verbal learning and verbal behavior*, 14, 638-647.

Ullman, M., Corkin, S., Coppola, M., Hickok, G., Growdon, J., Koroshetz, W., Pinker, S. (1997). A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by a procedural system. Journal of Cognitive Neuroscience 9 (2), 266-276.

Ullman, M. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, vol. 30, no. 1.

Vallar, G., Papagno, C. (2002). Neuropsychological impairments of verbal short-term memory. In: Baddeley, A., Kopleman, M., Wilson, B. (Eds). *Handbook of memory disorders*, 2nd ed, 249-270, Chichester: Wiley.

Van Dyke, J., McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language* 65, p.247-263.

_____. (2006). Retrieval interference in sentence comprehension. Journal of Memory and Language 55, 157–166.

Vigliocco, G., Franck, J. (1999). When sex and syntax go hand in hand: gender agreement in language production. *Journal of Memory and Language*, 40, 455-478.

Vigliocco, G., Vinson, D., Paganelli, F., Dworzynski, K. (2005). Grammatical gender effects on cognition: implications for language learning and language use. *Journal of Experimental Psychology*, vol. 134, no. 4, 501-520.

Villalva, A. (1994). *Unidades e Hierarquias nas Palavras do Português*. Lisboa: FCT e FCG.

Villalva, A. (2012). Palavras, que as há. Revista de Estudos Linguísticos, Belo Horizonte, vol. 20, n.2, 125-139.

Xiang, M., Dillon, B. and Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain & Language* 108, 40–55.

Appendix

Type	Word	Gender	Stereotype
Neuter	*Adolescente* (adolescent)	78% ambiguous	-0.08
Neuter	*Agente de seguros* (insurance agent)	71% ambiguous	0.08
Neuter	*Agente de viagens* (travel agent)	78% ambiguous	-0.08
Neuter	*Artista* (artist)	71% ambiguous	-0.41
Neuter	*Ativista político* (political activist)	50% ambiguous	0.19
Neuter	*Atleta* (athlete)	78% ambiguous	-0.14
Neuter	*Banhista* (bather)	71% ambiguous	0.02
Neuter	*Celebridade* (celebrity)	85% feminine	-0.52
Neuter	*Chefe de cozinha* (chef)	78% ambiguous	0.13
Neuter	*Ciclista* (cyclist)	57% ambiguous	0.02
Neuter	*Cliente* (client)	78% ambiguous	-0.08
Neuter	*Comediante* (comedian)	78% ambiguous	-0.02
Neuter	*Dentista* (dentist)	78% ambiguous	-0.52
Neuter	*Designer gráfico* (graphic designer)	64% ambiguous	-0.08
Neuter	*Diplomata* (diplomat)	71% ambiguous	0.19
Neuter	*Economista* (economist)	71% ambiguous	0.08
Neuter	*Estudante* (student)	64% ambiguous	-0.25
Neuter	*Fisioterapeuta* (physiotherapist)	71% ambiguous	-0.14
Neuter	*Geologista* (geologist)	71% ambiguous	-0.02
Neuter	*Gerente* (manager)	78% ambiguous	0.13
Neuter	*Ginasta* (gymnast)	78% ambiguous	-0.52
Neuter	*Ginecologista* (gynecologist)	85% ambiguous	-0.69
Neuter	*Guia de turismo* (tourism guide)	71% ambiguous	-0.19
Neuter	*Hóspede* (guest)	57% ambiguous	0.02
Neuter	*Informante* (informant)	71% ambiguous	0.02
Neuter	*Jornalista* (journalist)	85% ambiguous	-0.58
Neuter	*Líder* (leader)	85% ambiguous	0.37
Neuter	*Metereologista* (meteorologist)	78% ambiguous	-0.08
Neuter	*Neurologista* (neurologist)	78% ambiguous	-0.02
Neuter	*Paciente* (patient)	78% ambiguous	-0.02
Neuter	*Pediatra* (pediatrician)	71% ambiguous	-0.3
Neuter	Personal trainer	85% ambiguous	0.13
Neuter	*Presidente da empresa* (CEO)	85% ambiguous	0.13
Neuter	*Protestante* (protestant)	78% ambiguous	-0.14
Neuter	*Psiquiatra* (psychiatrist)	71% ambiguous	0.02
Neuter	*Radialista* (broadcaster)	78% ambiguous	-0.02

Neuter	*Repórter* (reporter)	71% ambiguous	-0.08
Neuter	*Romancista* (romanticist)	78% ambiguous	-0.41
Neuter	*Sem-teto* (homeless person)	71% ambiguous	0.13
Neuter	*Zoologista* (zoologist)	64% ambiguous	-0.02

Table 46: Results of Pre-tests for neutral bigenders

Type	Word	Gender	Stereotype
Mas Stereo	*Agente do FBI* (FBI agent)	71% ambiguous	0.24
Mas Stereo	*Analista de sistemas* (IT analist)	71% ambiguous	0.19
Mas Stereo	*Astronauta* (astronaut)	71% ambiguous	-0.02
Mas Stereo	*Bombeiro* (fireman)	71% masculine	0.63
Mas Stereo	*Cientista* (cientist)	78% ambiguous	0.08
Mas Stereo	Coronel	57% masculine	1.08
Mas Stereo	*Detetive* (detective)	71% ambiguous	0.41
Mas Stereo	*Eletricista* (electrician)	57% ambiguous	0.3
Mas Stereo	*Fã de corrida de carros* (car race fan)	64% ambiguous	0.3
Mas Stereo	*Frentista* (gas station attendant)	78% ambiguous	0.02
Mas Stereo	*Guarda* (guard)	50% ambiguous	0.47
Mas Stereo	*Motorista* (driver)	64% ambiguous	0.24
Mas Stereo	*Piloto* (pilot)	50% ambiguous	0.41
Mas Stereo	*Piloto de corrida* (race car driver)	57% ambiguous	0.91
Mas Stereo	*Policial* (police officer)	71% ambiguous	0.19
Mas Stereo	*Soldado* (soldier)	57% ambiguous	0.85
Mas Stereo	*Surfista* (surfist)	78% ambiguous	0.2
Mas Stereo	*Taxista* (taxi driver)	71% ambiguous	0.41
Mas Stereo	*Traficante* (drug dealer)	50% ambiguous	0.58

Table 47: Results of Pre-tests for masculine stereotypical bigenders

Type	Word	Gender	Stereotype
Fem Stereo	*Assistente de mágica* (magician assistant)	57% feminine	-1.32
Fem Stereo	*Assistente social* (social worker)	71% ambiguous	-0.64
Fem Stereo	*Babá* (babysitter)	57% feminine	-1.32
Fem Stereo	*Caixa* (cashier)	50% ambiguous	-0.69
Fem Stereo	*Colunista de auto-ajuda* (self-help columnist)	85% ambiguous	-0.58
Fem Stereo	*Design de interiores* (interior designer)	72% ambiguous	-0.47
Fem Stereo	*Diarista* (maid)	57% feminine	-1.14
Fem Stereo	*Esteticista* (beautician)	71% ambiguous	-0.64

Fem Stereo	*Florista* (florist)	78% ambiguous	-0.58
Fem Stereo	Manicure	78% ambiguous	-0.47
Fem Stereo	*Massagista* (massagist)	78% ambiguous	0.36
Fem Stereo	*Modelo* (model)	78% ambiguous	-0.69
Fem Stereo	*Nutricionista* (nutritionist)	78% ambiguous	-0.8
Fem Stereo	*Recepcionista* (receptionist)	78% ambiguous	-0.75
Fem Stereo	Socialite	57% ambiguous	-1.3
Fem Stereo	Stripper	50% ambiguous	-0.58
Fem Stereo	*Terapeuta* (therapist)	57% ambiguous	-0.64
Fem Stereo	*Terapeuta ocupacional* (occupational therapist)	85% ambiguous	-0.64
Fem Stereo	*Vítima de estupro* (rape victim)	92% feminine	-1.36

Table 48: Results of Pre-tests for feminine stereotypical bigenders

Type	Word	Gender	Stereotype
Mas Epi	*Algoz* (executioner)	57% masculine	0.35
Mas Epi	*Anjo* (angel)	92% masculine	0.3
Mas Epi	*Astro de cinema* (movie star)	78% masculine	0.69
Mas Epi	*Bebê* (baby)	57% masculine	-0.02
Mas Epi	*Bicho* (animal)	92% masculine	0.47
Mas Epi	*Boia-fria* (farmer worker)	64% ambiguous	0.13
Mas Epi	*Carrasco* (executioner)	85% masculine	0.97
Mas Epi	*Cônjuge* (spouse)	64% ambiguous	0.02
Mas Epi	*Dedo-duro* (snitch)	64% ambiguous	-0.08
Mas Epi	*Defunto* (corpse)	78% masculine	0.47
Mas Epi	*Ente* (entity)	64% masculine	0.02
Mas Epi	*Gênio* (genius)	78% masculine	0.69
Mas Epi	*Ídolo* (idol)	92% masculine	0.63
Mas Epi	*Indivíduo* (individual)	92% masculine	0.19
Mas Epi	*Membro* (member)	92% masculine	0.63
Mas Epi	*Monstro* (monster)	100% masculine	0.91
Mas Epi	*Neném* (baby)	64% ambiguous	-0.25
Mas Epi	*Ser* (being)	92% masculine	0.19

Table 49: Results of Pre-tests for masculine grammatical gender

Type	Word	Gender	Stereotype
Fem Epi	*Criança* (child)	85% feminine	-0.58
Fem Epi	*Criatura* (creature)	92% feminine	-0.54
Fem Epi	*Estrela de cinema*	85% feminine	-0.50

	(movie star)		
Fem Epi	*Pessoa* (person)	92% feminine	-0.47
Fem Epi	*Testemunha* (witness)	57% feminine	-0.58
Fem Epi	*Visita* (guest)	92% feminine	-0.64
Fem Epi	*Vítima* (victim)	92% feminine	-0.58

Table 50: Results of Pre-tests for feminine grammatical gender

[i] Sample of the materials of Lago (2014)

Experiment in Spanish:

Experiment in Spanish:

Gram, sg attractor: La nota que la chica **va** a escribir en la clase alegrará a su amiga.

Gram, sg attractor: La nota que la chica **va** a escribir en la clase alegrará a su amiga.

(The note that the girl are going to write during class will cheer her friend up.)

Gram, pl attractor: Las notas que la chica **va** a escribir en la clase alegrará a su amiga.

(The notes that the girl are going to write during class will cheer her friend up.)

Ungram, sg attractor: *La nota que la chica **van** a escribir en la clase alegrará a su amiga.

(The note that the girl are going to write during class will cheer her friend up.)

Ungram, pl attractor: * Las notas que la chica **van** a escribir en la clase alegrará a su amiga.

(The notes that the girl are going to write during class will cheer her friend up.)

Experiment in English:

Gram, sg attractor: The musician that the reviewer **was** highly praising last week will probably win a Grammy.

Gram, pl attractor: The musicians that the reviewer **was** highly praising last week will probably win a Grammy.

Ungram, sg attractor: *The musician that the reviewer **were** highly praising last week will probably win a Grammy.

Ungram, pl attractor: *The musicians that the reviewer **were** highly praising last week will probably win a Grammy.