



THE EFFECT OF GENDER AGREEMENT MISMATCHES ON INTERSENTENTIAL ANAPHORIC PRONOUNS

O EFEITO DAS INCONGRUÊNCIAS DE CONCORDÂNCIA DE GÊNERO EM PRONOMES ANAFÓRICOS INTERSENTENCIAIS

Eduardo Correa Soares¹
Mailce Borges Mota²

Abstract: This article focuses on the effect of gender agreement mismatches between personal pronouns and their antecedents across sentences. In two acceptability experiments, we test whether acceptability of gender agreement violations on animated nouns may be modulated by grammatical and contextual features of the antecedents of personal pronouns. In the first experiment, we manipulated the “specificity” feature of the antecedent in order to make the antecedent refer either to the class of individuals or to a specific referent. In the second experiment, we used stereotypically male or female proper names to test whether grammatical gender mismatches between personal pronouns and bigender nouns could be attenuated. Although the first experiment showed an effect explainable purely by grammatical factors, against many theories of “semantic” agreement, the results of the second experiment suggest that both the grammatical and the contextual features of the antecedent are computed when speakers evaluate agreement relations between personal pronouns and their antecedents.

Keywords: Gender agreement, acceptability judgements, morphosyntax, pragmatics.

Resumo: Este artigo enfoca o efeito da incongruência da concordância de gênero entre pronomes pessoais e seus antecedentes entre sentenças. Em dois experimentos de aceitabilidade, testa-se se a aceitabilidade de violações de concordância de gênero em substantivos animados pode ser modulada por características gramaticais e contextuais dos antecedentes de pronomes pessoais. No primeiro experimento, manipula-se a “especificidade” do antecedente para fazer o antecedente referir-se à classe de indivíduos ou a um referente específico. No segundo experimento, utilizando nomes próprios estereotipicamente masculinos ou femininos, testa-se se a incongruência gramatical de gênero entre pronomes pessoais e substantivos sobrecomuns e comuns de dois gêneros pode ser atenuada. Embora o primeiro experimento tenha mostrado um efeito explicável puramente por fatores gramaticais, contra muitas teorias da concordância “semântica”, os resultados do segundo experimento sugerem que as características gramaticais e contextuais do antecedente são computadas em paralelo quando os falantes avaliam relações de concordância entre pronomes pessoais e seus antecedentes.

Palavras-chave: Concordância de gênero, julgamentos de aceitabilidade, morfossintaxe, pragmática.

1. INTRODUCTION

This article focuses on the effect of gender agreement mismatches between pronouns and their antecedents across sentences. In two experiments with Brazilian Portuguese native speakers, we used a paradigm with agreement violations to individually test grammatical and contextual properties of agreement relations between pronouns and their antecedents. Our results show that, against previous empirical observations across languages, gender agreement between pronouns and their antecedents is subject to both grammatical and contextual gender features. This is so because the acceptability of pronouns linked to an antecedent is influenced by its contextual and grammatical features in the previous sentences.

¹ Pesquisador na Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brasil. soares_ec@yahoo.com.br
Orcid: <https://orcid.org/0000-0002-4526-3299>

² Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brasil. mailce.mota@ufsc.br
Orcid: <https://orcid.org/0000-0002-8674-2480>

Gender agreement seems to be a strong constraint on the choice of anaphoric pronouns: as for animate masculine nouns, only the masculine pronoun *ele* “he” is generally accepted; on the other hand, only the feminine pronoun *ela* “she” is usually acceptable when referring to a feminine animate noun, as in (1) below. Notice that this restriction appears to apply to any anaphoric relation regardless of reference of the common noun being specified or not, as illustrated by example (2) below.

- (1) Mas a garota₁ é novinha. *Ele/Ela₁ ainda faz faculdade...
“But the girl₁ is young. *He/She₁ is still an undergrad.”
- (2) ...no dia seguinte você pegar um ônibus e ir pra casa da pessoa₁, e, pô, você pensar que queria ficar mais tempo com ??ele/ela₁...
“...in the next day, you get on a bus and go to someone₁’s house [literally, ‘the person’s house’], and, so, you think you want to stay more time with ??him/her₁...”

(NURC-RJ – Interview_003ac)

Agreement is considered a central phenomenon in native speakers’ competence (see CORBETT, 2006; JOHNSON, 2014, among others). In addition, agreement has also been recently studied by several psycholinguists because it might shed light on aspects of human cognition (LEWIS; VASISHTH, 2005; SLEVC; MARTIN, 2016; PARKER et al., 2017, among others). In this vein, agreement between anaphoric expressions and their antecedents is a key piece of evidence for understanding language and cognitive architecture and how this architecture shapes language processing.

In this paper, we investigate the properties of intersentential anaphoric agreement. We took a romance language (Brazilian Portuguese) as our sample because in this language all nouns are either lexically or morphosyntactically marked for gender and, interestingly, some morphosyntactically gender-marked nouns may refer to both male and female individuals. Specifically, we built two experiments to try to disentangle the effects of grammatical and contextual agreement between subject pronouns and their antecedents. In the first experiment, we used sentences starting with specific and non-specific masculine nouns retrieved by either masculine or feminine pronouns. The issue under investigation was whether nonspecific nouns, which may contextually refer to both biological genders, would be more acceptable when retrieved by a feminine pronoun than their specific counterparts. We found a strong effect of grammatical agreement, which was not modulated by specificity. However, given previous results in the literature, we suspected that only grammatical agreement is not sufficient to account for agreement relations between pronouns and their antecedents and that our first experiment was not drawing the full picture. In order to investigate what the preferential option is when the morphosyntactic gender of animate nouns is not convergent with the biological gender of the referent, we carried out a second experiment. In our second experiment, we used thirty-six animate common nouns that may refer to both biological genders, half of them referring to people and the other half referring to animals (“epicenes”). We manipulated the contextual gender by attributing proper names to the referents of these common nouns, which may stereotypically refer to male or female individuals. Our results show that the contribution of morphosyntactic and contextual gender is equally important for the acceptability of the sentences in which personal pronouns refer back to their antecedents.

2. THEORETICAL BACKGROUND

Several hypotheses about the relation of agreement divide the analysis between what would be predicted by linguistic representations and what would be in charge of general multi-purpose cognitive processes, such as working memory (see SLEVC; MARTIN, 2016 and references cited there). We will call these two approaches Representational Models and Retrieval Models, respectively. Each one of them accounts for agreement mismatches in different ways. However, there are few studies analyzing how to reconcile the interface between contextual and grammatical agreement and the theory of general cognitive processes, *i. e.*, a theory that combines Representational Models and Retrieval Models. We briefly review these approaches in the next

sections in order to build a broad picture of these theories and advance our proposal, which combines these two types of models.

2.1 Overview of Agreement and Gender

Agreement is generally taken to be a grammatical relation between two or more elements that co-vary (STEELE, 1978, p. 610). Matthews (1981, p. 247), for example, states that “[agreement] ... is usually described as a relation between words that share a morphosyntactic feature”. Corbett (2006, p. 1) draws a broader line to cover all sorts of phenomena usually called agreement, whose key notion is “displacement of information”. According to him, an agreement relation is characterized by one word displaying information from the features of another word. One central observation that has drawn considerable attention in the literature is the fact that semantic features can also fire agreement relations. The so-called “semantic agreement”, or the interface between morphosyntax, semantics and pragmatics, has been studied by many authors, since this phenomenon seems pervasive across languages (e.g. COMRIE, 1975; CORBETT, 1979, 2006; WECHSLER; ZLATIC, 2003, among others).

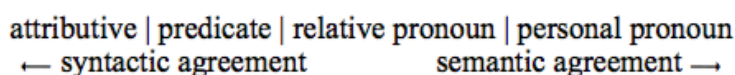
Furthermore, agreement has recently become a tangential point of formal and functional approaches in linguistics (JOHNSON, 2014). The phenomena involving agreement is studied within grammatical formal models (e.g. percolating, copying, matching, sharing, inheriting, etc., agreement features) and/or within functional models of language production and processing (as a part of the cognitive resources and processes and the organization and encoding of grammatical and content information within the brain). Some theories postulate a “division of labor” between memory and linguistic representation with respect to agreement (see SLEVC; MARTIN, 2016 and research cited there). However, much of this research is restricted to grammatical agreement, not taking into account that the phenomenon of agreement is at the interface between morphosyntax, semantics and pragmatics. We could group proposals found in the literature into two main hypotheses for the treatment of agreement: the Representational Models and the Retrieval Models. Although there are different implementations of these ideas, Representational Models are based on the computation of linguistic features that percolate throughout the linguistic representation (see, for example, VIGLIOCCO; NICOL, 1998 and work with their colleagues). The Retrieval Models are based on information that is stored in memory and its limitations. “Clues” are used to reactivate in memory the pertinent referent, whose agreement features coincide with the greater number of “clues” associated with it (see LEWIS; VASISHTH, 2005; VAN DYKE; MCELREE, 2006; 2011; PATIL et al., 2016; PARKER et al., 2017). Before we move to the next section, in which we present an overview of these two hypotheses, let us briefly introduce the notions of “gender” and “gender agreement”.

Unlike number inflection and agreement, gender inflection and gender agreement are taken to be phenomena more related to grammar than to meaning. According to Corbett (2006), the crosslinguistic existence of different genders that do not correspond to a specific semantic content associated with each class shows that the association between semantic and syntactic gender is not trivial: in Romance languages (except Romanian), for example, inanimate referents are necessarily masculine or feminine (CORBETT, 1994); in Russian, there are three genders (masculine, feminine and neutral), but there is no evident correspondence of inanimate with neuter and between masculine / feminine and the respective semantic (or biological) gender of referents (CORBETT, 1979). The conclusion reached by many studies is that gender is an arbitrary morphosyntactic classification, without a well-defined semantic counterpart (see HOCKETT, 1958; DIXON, 1968; FRANCESCHINA, 2005, among many others). Given these notions, it would be expected that gender agreement would be much more accountable by grammatical feature sharing than a phenomenon subject to semantic or contextual displacement of information. As shown in the next section, this prediction goes against many theories that propose that personal pronouns are subject to semantic agreement rather than grammatical agreement.

2.2. Representational Models

In representational terms, different types of agreement are identified in many languages (CORBETT, 1991; POLLARD; SAG, 1994; KATHOL, 1999; JOHNSON, 2014). Some of these authors postulate an interrelation between different components of the grammar to explain agreement relations or a predominance of one module of the grammar on another for each different agreement relation. Corbett (1979, 1991, 2006), for example, proposes that semantic agreement is more likely with some syntactic categories than others, according to an Agreement Hierarchy (AH): the predominance of semantic agreement increases monotonically rightward as the agreeing form is more nominal.

Figure 1: Agreement Hierarchy



From a typological standpoint, Croft (2001, 2012) and Haspelmath (2012) put forth a proposal according to which agreement and coindexing (as in the case of pronouns) are the same linguistic phenomenon (see also DOWTY; JACOBSON, 1988). They call into question whether there should really be any distinction between pronouns and agreement markers. In their view, agreement may be reduced to coindexing because pieces of meaning are often expressed in more than one form in a single sentence, *i. e.*, there is nothing special in two agreeing elements expressing a single meaning. A different typological perspective is entertained by Johnson (2014). Studying Latin, Ancient Greek and Albanian based on corpora and elicitation tasks, she observes that the choice of one type of agreement over the other is strongly dependent on contextual information and puts forth a performance-based view of agreement. Under her approach, two mechanisms are crucial in agreement relations: “resolution”, as a sort of feature assignment, and “partial agreement”, as an avoidance strategy. Her proposal is built upon existing machinery in the grammar (a unification grammar), but the result are constraints that are cumulative of the grammatical, semantic and pragmatic information in context.

In unification grammars, such as the HPSG, the central concept of agreement theory is the “index”, which is a linguistic entity linked to an object in the interpretation domain. Using this theory, Pollard & Sag (1994), for example, propose that, in English, the agreement between determinant and noun is essentially morphosyntactic, whereas the agreement between pronoun and antecedent is a reflection of semantic and pragmatic index agreement: co-indexing. However, they distinguish between the index *per se* and the properties of an index with respect to how it refers to an entity of the world, which is within the “context” feature (equivalent to pragmatics). Their mechanism to account for agreement relations is selection: a masculine plural adjective, for instance, selects (*i. e.*, imposes restrictions on the index of) a masculine plural noun. Kathol (1999), Wechsler and Zlatic (2000), Villavicencio et al. (2005), and An and Abeillé (2017) further refine this idea by studying numerous phenomena in different languages and proposing several features and restrictions to account for different patterns for each pair of categories involved in agreement relations. What must be retained from these analyses is that (i) they are multilayered, *i. e.* different components of the linguistic representation are computed in agreement relations; (ii) as in the AH, some pairs of categories are taken to be more subject to a certain type of agreement than to another; and (iii) to the best of our knowledge, no approach (excepting Johnson 2014) proposes graduality in the effects of agreement mismatches, that is, once an agreement relation is violated, the sentence is supposed to fail to be accepted.

Corbett (1991) also observes that the associations between grammatical gender and a semantic counterpart is complex: in French, for example, *les Américains* “the Americans.MASC”, which is a grammatically masculine noun, may refer to a group of people composed by male and female individuals, while *les Américaines* “the Americans.FEM” may refer to a group composed of only female individuals. For this reason, he proposes that masculine gender is the unmarked, default gender in French. An and Abeillé (2017) further argue that gender conflicts in coordinated

structures, such as *Le garçon et la fille sont compétents/*compétentes*. “The boy and the girl are competent.MASC/*FEM.PL”, are usually solved by gender resolution agreement, in which case the default value is preferred: in French, masculine adjective. Alves (2019) makes a similar point on Portuguese and Spanish by arguing that masculine gender neutralizes semantic gender conflicts. Here, we are not going to discuss plural nouns because they involve a semantic description of “plural”, which is beyond the scope of this paper. However, we assume that masculine is the unmarked gender in Portuguese for the cases in which, having the two grammatical genders available for a given lexical item, the speaker is forced to choose a grammatical gender, even if it denotes different biological genders.

These different levels of analysis may have processing consequences, as will be studied in this article. Johnson (2014) suggests that even if agreement has some formal component (such as feature sharing), it is built in many cases “on the fly”. In the psycholinguistic literature, some explanations to account for “on the fly” properties of agreement were put forth. There are many proposals and reviewing all of them would be beyond the scope of this paper. In the next section, we briefly give an overview of some of them in order to build the ground for our discussion.

2.3. Retrieval Models

Psycholinguists have long observed that agreement properties, specially the so-called “attraction errors”, may be a window into the way speakers process language. Based upon the idea that for language processing to take place some information must be temporarily kept in memory (LEWIS et al., 2006), researchers proposed that agreement relations, as well as coreference resolution, use a multi-purpose cognitive processing device to temporarily store some information up to the point when integration takes place. “Attraction errors” are evidence for the use of this temporary memory in agreement relations, since as memory is subject to cognitive burden and other sorts of interference, effects of linear distance and other performance factors may lead speakers to produce agreement deviances or accept them fine.

To account for these observations, researchers proposed memory-based “retrieval” models. One of these models is called Content-Addressable Memory (CAM), according to which the antecedent for an agreement relation is retrieved by parallel search on the basis of “cues” generated by the target (MCELREE, 2000; MCELREE et al., 2003; VAN DYKE; MCELREE, 2006, 2011). Interferences may hinder the resumption of certain entities, especially when their traits and distractors converge (see LEWIS; VASISHTH, 2005; PARKER et al., 2017, among others). Solving long-distance dependencies, such as agreement of personal pronouns and their antecedents, is a calculus, incrementally augmented as the sentence unfolds, which is subject to grammatical, semantic and pragmatic constraints.

Parker et al. (2017), however, point out that the cue-based theory of retrieval lacks a clear theory of cues: the account offered by Martin and McElree (2008), for example, is not explicit on how grammatical and interpretive constraints are mapped onto retrieval cues. Although sketched implementations of retrieval cues are many in psycholinguistic theories, we can broadly group them into two: (i) equal combination, in which all cues are combined equally at retrieval, as in other models of memory retrieval (CLARK; GRONLUND, 1996); and (ii) non-uniform mapping, in which only a subset of features is used as retrieval cues (DILLON, 2011; DILLON et al., 2013) or contribute in a weighted cue-combinatorics scheme (PARKER; PHILLIPS, 2014; 2017). Given empirical evidence obtained so far, it is very unlikely that all linguistic information is used together with the same weights in retrieval (see PARKER et al., 2017 and references there). Our experiments, reported in section 3, contribute to the discussion of which linguistic information is stored as cue in memory and may be used for retrieval. Before moving to our data, previous research on Brazilian Portuguese gender agreement in personal pronouns is summarized in what follows.

2.4. Previous Experimental Studies on Brazilian Portuguese

Elaborating on a study on Italian by Cacciari et al. (1997), Corrêa (2001) studied the processing of gender agreement and the comprehension of pronouns in different discourse

contexts. In two self-paced reading experiments, she manipulated the semantic-pragmatic gender of bigender common nouns by using adjuncts stereotypically attributed to a male or to a female person or to both of them, such as *A testemunha de meia-calça preta* “the witness in black stocks” (female referent) / *de barba preta* “with a black beard” (male referent) / *de roupas claras* “in light clothes” (neuter). In her Experiment 1, a first, context sentence, in which the bigender common noun was presented, was followed by a second, target sentence starting with a pronoun grammatically agreeing with that first bigender common noun. The pronoun was, thus, incongruent with contextual gender of one of the interpretations produced by the adjunct. In her Experiment 2, a sentence was inserted between the antecedent (the bigender common noun) and the pronoun. The intention was to make it harder to recover the grammatical information of the bigender noun. In both experiments, she found a significant difference in the reading times by the end of the sentence that started with the pronoun, with the contextually incongruent gender being read slower.

Lawall, Maia & Amaral (2012) investigated the processing of matching and mismatching pronouns with bigender common nouns in intrasentential contexts in native BP speakers, heritage speakers and advanced BP learners through a self-paced reading task. Their experiment presented six conditions: the first four were sentences in which the first clause had grammatically feminine or masculine common nouns that could refer to both biological genders, followed by another clause that started with personal pronouns, with either gender, such as *a vítima* “the victim.FEM” ... *ela/ele* “she/he” and *o bebê* “the baby.MASC”... *ela/ele* “she/he”. The other two sentences were sentences that started with bigender common nouns, whose grammatical and semantic gender was specified by the determiner, such as *a assistente* “the.FEM assistant”... *ela* “she” and *o assistente* “the.MASC assistant”... *ele* “he”. For the native-speakers group, significant higher reaction times (RTs) were found in the personal pronouns’ region in a comparison between the conditions grammatically matched and mismatched with common nouns that referred to both biological sexes. Lawall, Maia & Amaral (2012) attribute these higher RTs to “feature sharing” between the grammatical features and the semantic features of the antecedent. In section 4, we approach these results in a different way in the light of the results we report in section 3.

Finally, Alves (2014) tested coreferential personal pronoun whose antecedents were either bigender plural common nouns, such as *as estudantes* “students.MASC/FEM”, or nouns that could refer to both biological genders, such as *as indivíduos* “individuals.MASC”, which were also inflected in the plural form. She used stereotypical contexts to manipulate the contextual gender of the referents, such as a stereotypical predicate of female individuals “looking for a prenatal exam...” vs. a neuter one “looking for better jobs”. Alves (2014) found significantly higher RTs in grammatical gender mismatched conditions, such as *as vítimas...* *elas* “the victims.FEM... they.MASC”) and a secondary effect of context. According to Alves (2014), these results suggest a hierarchy of factors used in online pronominal processing in Brazilian Portuguese: (i) the most important factor is the grammatical gender agreement and (ii) contextual gender match has a secondary role in the anaphora resolution³.

3. EXPERIMENTAL EVIDENCE

In this section, we report two acceptability experiments carried out in order to disentangle the effects of grammatical and contextual gender on the agreement relations between pronouns and their antecedents. The first experiment was designed to test whether the predictions made by the Agreement Hierarchy could be observed in a context where the semantic content of the antecedent is weakened: with nonspecific antecedents, given that personal pronouns are at the rightmost position of the hierarchy, *i. e.* they are highly driven by the semantic agreement, we would expect that either gender would be equally acceptable. That was not what we found: purely

³ Leitão, Peixoto & Santos (2008), Alves (2016), Alves (2017) and Alves (2019) also carried out experiments manipulating gender. However, the most crucial manipulations and results are related to Principle B (CHOMSKY 1981), according to which a pronoun may not have its antecedent in a certain syntactic intrasentential domain. Since we are interested in the use of pronouns recovering antecedents in the previous sentence, their results are not directly comparable to ours.

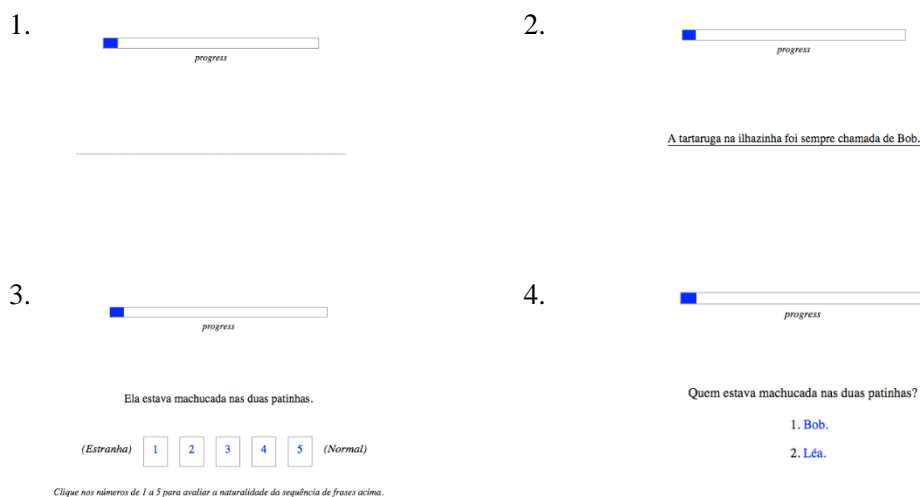
grammatical disagreement with nonspecific antecedents was judged as low as with specific antecedents. We then carried out another experiment to test whether agreement relations between pronouns and their antecedents were driven purely by grammatical features: we used nouns that can denote referents with both biological genders combined with stereotypical proper nouns in order to test whether this contextual gender (given by the proper noun) might render the sequence more acceptable. Our results show that both the grammatical and the contextual gender exert effects on the acceptability of the sentences exactly at the same degree: violating grammatical agreement is essentially not different from violating contextual agreement with respect to the acceptability of the sentences in which the pronoun is used.

3.1. Method

This experiment was approved by the Ethical Board of UFSC (trial number: CAAE 89160418.2.0000.0121). Participants were invited to take part in our experiment via email and social networks. All participants voluntarily participated in the experiments on the IbexFarm platform (DRUMMOND, 2014). They received a link through which they were redirected to the first page of our experiment – the Consent Form. They filled in a basic anonymous information form and read the instructions to participate. They had three trials to practice before the real experiment started.

Each trial started with a horizontal bar (screen 1 in Figure 2). Participants were asked to press “space bar” on their keyboard to see the first sentence of the pair (screen 2 in Figure 2). Pressing “space bar” again led them to the second sentence, when they had to judge how acceptable it was in the context on a Likert scale from 1 to 5 (screen 3 in Figure 2). They were instructed to use the full scale according to how natural “Normal” or strange “Estranha” the sentence was. Finally, after judging the second sentence, they had to answer a forced-choice interpretation question with two possible answers (screen 4 in Figure 2).

Figure 2: Screen Samples



For each experiment, the thirty-six target-items were randomized with ninety-six distractors. The distractors showed homophonous words with different meanings depending on gender (*o/a bandeirinha* “the.MASC/the.FEM auxiliary-judge.MASC/little-flag.FEM”) and dubiously masculine or feminine inanimate nouns (*o/a alface* “the.MASC/the.FEM lettuce”). All target-items and distractors had the same number of words. Each participant was presented with one single condition of each experimental item – a Latin-square design crossing participants and items (BAYEEN et al., 2008). The sentences were fully randomized in a way that no participant judged the same list (that is, the same conditions of the same items in the same order). Three control sentences that violate strong grammatical constraints were inserted in the end of the experiment, in order to assure that participants were attentive until the end and to avoid ceiling effects.

3.1. Experiment 1

Experiment 1 was designed to test the predictions of the Agreement Hierarchy regarding personal pronouns and the conjectures on the lack of semantic counterpart in morphosyntactic gender. As mentioned before, the predictions drawn from the AH seem counterintuitive if gender is an exclusively morphosyntactic noun classifier. We took advantage of the idea that agreement is a “displacement of information” to design this experiment. We used a paradigm with gender agreement violations between pronouns and their antecedents, the last of which are bigender nouns whose gender is visible by the syntactic context (such as *o estudante* “the.MASC student”). These antecedents were either specific or nonspecific. This manipulation made the semantic gender visible or not, because nonspecific antecedents refer to the class defined by the noun, not to a specific referent. Our results suggest that regardless of having a visible semantic gender, the morphosyntactic information exerts effects on the acceptability of the pronouns in the sentences.

3.1.1. Rationale

As mentioned before, in this experiment, we built pairs of sentences to test the pronoun-antecedent agreement relations across sentences. The aim of this experiment was to test whether pronoun-antecedent agreement relations is more semantic, since according to the AH, personal pronouns are at the rightmost point of semantic agreement property. On the other hand, many researchers, based on crosslinguistic data, conjectured that gender inflection and agreement are solely grammatical phenomena. Both claims together make contradictory predictions with respect to gender agreement relations between personal pronouns and their antecedents.

We manipulated specificity of the antecedent in order to make the semantic gender salient or not to isolate grammatical gender and agreement and to test these claims. In the first sentence, we made an antecedent available and salient: a bigender common noun, whose gender is marked by the determiner immediately before it. The crucial condition is the nonspecific one, since nonspecific antecedents refer to the class of individuals and thus, do not have a semantic or contextual gender. In such case, if the AH is correct, choosing an agreeing or a disagreeing pronoun should not exert any effect on the acceptability of the sentences. That should contrast with the specific condition, in which a disagreeing pronoun should be less acceptable than an agreeing pronoun. On the other hand, if the conjectures on gender inflection and agreement being a more grammatical phenomena are right, regardless of specificity disagreeing pronouns should be less acceptable than agreeing pronouns.

3.1.2. Stimuli

Thirty-six items were created for this experiment. They started with a grammatically masculine determiner followed by a bigender common noun. They were counterbalanced with filler items starting with a grammatically feminine determiner followed by bigender common nouns with a similar structure. A total of 144 sentences were built displaying the two experimental factors (*antecedent* and *agreement*), each one with two levels, in a two-by-two design, as shown

in Table 01. The first sentence of the pair (hereafter, *context*) was used to manipulate Factor *antecedent*: in the *Nonspecific* condition, the determiner is indefinite and the verb is modal, so that the interpretation of the sentence becomes generic and the reference of the antecedent is nonspecific (denoting the class of individuals); in the *Specific* condition, the determiner is definite and the verb is in the past tense, so that the interpretation of the sentence becomes episodic and the reference of the antecedent is specific (denoting a specified individual). The second sentence of the pair (hereafter, *target*) was used to manipulate the second factor: in the *Agrees* condition, the pronoun in the target sentence agrees with the morphosyntactic gender of antecedent in the context sentence; in the *Disagrees* condition, the gender of the pronoun in the target sentence is different from the morphosyntactic gender of the antecedent in the context sentence. Each experimental item was followed by an interpretation question in order to assure that participants were attentive to the experiment and that they were interpreting the pronoun as co-referential with the relevant noun, as those of the sample in Table 01.

Table 01: Stimuli – Experiment 1

antecedent	agreement	Context	target	question
<i>Nonspecific</i>	<i>Agrees</i>	<i>Um agente do FBI</i> não pode jamais ser identificado. “An.MASC FBI agent must never be identified.”	<i>Ele</i> não foi descoberto pelo inimigo. “He was not discovered by the enemy.”	Quem não tinha sido descoberto pelo inimigo? “Who was not discovered by the enemy?”
<i>Nonspecific</i>	<i>Disagrees</i>	<i>Um agente do FBI</i> não pode jamais ser identificado. “An.MASC FBI agent must never be identified.”	<i>Ela</i> não foi descoberta pelo inimigo. “She was not discovered by the enemy.”	Quem não tinha sido descoberta pelo inimigo? “Who was not discovered by the enemy?”
<i>Specific</i>	<i>Agrees</i>	<i>O agente do FBI</i> não tinha jamais sido identificado. “The.MASC FBI agent had never been identified.”	<i>Ele</i> não foi descoberto pelo inimigo. “He was not discovered by the enemy.”	Quem não tinha sido descoberto pelo inimigo? “Who was not discovered by the enemy?”
<i>Specific</i>	<i>Disagrees</i>	<i>O agente do FBI</i> não tinha jamais sido identificado. “The.MASC FBI agent had never been identified.”	<i>Ela</i> não foi descoberta pelo inimigo. “She was not discovered by the enemy.”	Quem não tinha sido descoberta pelo inimigo? “Who was not discovered by the enemy?”

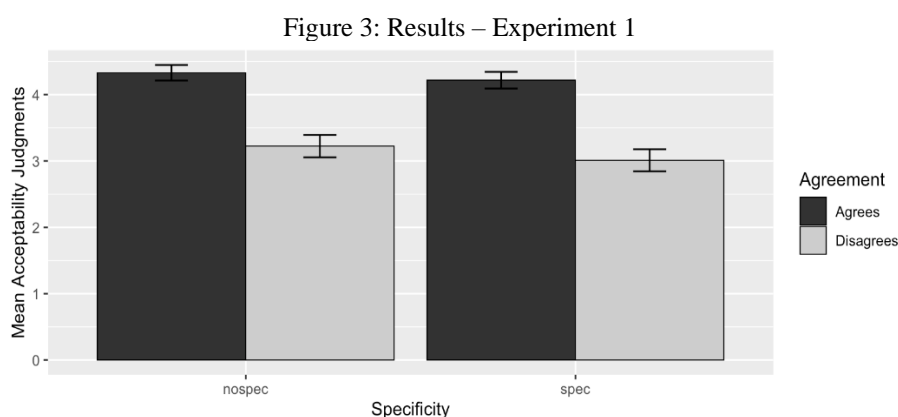
3.1.3 Predictions

Following the literature on gender inflection and agreement, the empirical predictions with respect to the acceptability judgments were the following. (i) If participants rely on the semantic features of the antecedent to fire agreement relations, when the antecedent is nonspecific, personal pronouns agreeing or disagreeing with these antecedents will be equally acceptable; conversely, when the antecedent is specific, personal pronouns disagreeing with the gender of the antecedent will be judged worse than personal pronouns agreeing with the antecedent: in such case, the effect of the interaction between factors should come up; or (ii) if participants take the grammatical features of the antecedent as decisive for the agreement relations between pronouns and their antecedents, personal pronouns disagreeing with the gender of the antecedent will be judged worse than personal pronouns agreeing with the antecedent regardless of the specificity of the antecedent: in such case, only the main effect of Factor Agreement should come up. Main Factor Specificity is orthogonal to our research questions here.

3.1.4 Results

Forty-five (twenty-five female) Brazilian Portuguese speakers living in the southern area of the country participated in this experiment. From this group, ten participants (seven female and three male) were excluded from the study because they informed that they were studying linguistics or language-related subjects (translation, literature or linguistics). Their data, therefore, was not analyzed. All data from the remaining participants ($N = 35$), who judged control unacceptable sentences lower than two and a half in average, were analyzed. Ninety-nine percent of the answers to the interpretation questions were correct. Acceptability judgments from trials with incorrect answers were thus excluded from data set.

Figure 3 summarizes the results of the experiment. Participants judged the sentences in which the personal pronouns agreed with the antecedent better (average: 4.27) than those sentences in which the pronoun disagreed with its antecedent's grammatical gender (average: 3.11). Being specific or nonspecific does not increase the acceptability of either condition.



Acceptability judgments were entered into a log-linear mixed-effects model analyses with random slopes (BARR et al., 2013). Although we kept the dummy coding for factor Agreement with level Agrees as the baseline 0 and Disagrees as 1, we set factor Specificity to be contrasted to average response by sum coding (SCHAD et al., 2019; BREHM; ALDAY, 2020): Non-specific -0.5 and Specific 0.5. The log-linear mixed-effects model was fitted by REML and t-tests used Satterthwaite's method (KUZNETSOVA et al., 2016).

Table 02: Log-linear mixed-effects model of Experiment 1

<i>Fixed Effects</i>	<i>Estimate</i>	<i>S. E.</i>	<i>T-val</i>	<i>Pr(> t)</i>
<i>Intercept</i>	4.293	0.106	40.38	< 2e-16
<i>Specificity</i>	-0.098	0.090	-1.08	0.287
<i>Agreement</i>	-1.154	0.199	-5.78	1.6e-06
<i>Specificity:Agreement</i>	-0.092	0.114	-0.81	0.423

As shown, the only factor modulating the acceptability judgements is Agreement: when the pronoun disagrees with its antecedent, regardless of being specific or non-specific, the acceptability of the sentence falls, which is noticed in the negative coefficient in Table 02.

3.1.4 Discussion

As predicted by the general crosslinguistic empirical observations, results of Experiment 1 show that gender agreement is strongly driven by the grammatical features of pronouns and their antecedents. Against the prediction drawn from the AH, gender agreement violations seem to be perceived by naïve speakers independently from semantics. Our results are also surprising in the context of research on Brazilian Portuguese, since previous research has consistently found effects of semantic and contextual gender on the processing of coreferential personal pronouns. There are two possible explanations for the conflict between our results and previously reported

results: (i) online processing methods (such as self-paced reading) reveal different aspects of linguistic knowledge that are different from what may be observed in acceptability judgements; and (ii) what our results show is that grammatical agreement is independent from semantic or contextual gender, but they do not suggest that semantic and contextual gender do not exert any effect on the acceptability of agreement relation between pronouns and their antecedents. Experiment 2 was designed in order to test (ii). We briefly address (i) in section 4.

3.2. Experiment 2

In Experiment 2 we aimed at disentangling the effects of morphosyntactic (grammatical) and semantic-pragmatic (contextual) gender agreement violations between pronouns and their antecedents. We used 36 animate common nouns that may refer to both biological genders, 21 of them referring to people and the others referring to animals (“epicenes”). Our results show that the contribution of grammatical and contextual gender is equally important for the acceptability of the sentences.

3.2.1. Rationale

In this experiment, we built pairs of sentences to test the pronoun-antecedent agreement relation across sentences. In the first sentence, stereotypically male or female proper names are used to produce contextual divergence or convergence with the grammatically masculine or feminine common noun referring to an animate entity (either a person or an animal). This strategy is used to disentangle grammatical and contextual gender in order to test the nature of gender agreement. As mentioned above, agreement is essentially a grammatical phenomenon related to the morphosyntactic category of common nouns. Other researchers claim that agreement is essentially a semantic and pragmatic phenomenon related to the conceptualization of common nouns. Having a mismatch between the grammatical and the semantic-pragmatic gender in the antecedents enables us to test these hypotheses.

3.2.2. Stimuli

Thirty-six items were created for this experiment. Half of them started with a grammatically masculine common noun. This counterbalance was crossed with another counterbalance between nouns referring to a person and those referring to animals (“epicenos”). A total of 144 sentences were built displaying the two experimental factors (*antecedent* and *agreement*), as shown in Table 03. The first sentence of the pair (hereafter, *context*) was used to manipulate Factor *antecedent*: in the *Reference Level*, the morphosyntactic gender of the subject was convergent with the contextual gender associated to the stereotypically masculine or feminine proper noun, which is a predicative; in the *Antecedent Mismatch* condition, the morphosyntactic gender of the subject was different from the contextual gender of the predicative. The second sentence of the pair (hereafter, *target*) was used to manipulate the second factor: in the *Agrees* condition, the pronoun in the target sentence agrees with the morphosyntactic and contextual gender of the antecedent in the context sentence; in the *Disagrees* condition, the pronoun in the target sentence is different from the morphosyntactic and contextual gender of the antecedent in the context sentence; in the *Grammatical Agreement* condition, the pronoun in the target sentence converges with only the morphosyntactic gender of the antecedent in the context sentence; and in the *Contextual Agreement*, the pronoun in the target sentence matches only the contextual gender of the antecedent in the context sentence. Each experimental item was followed by an interpretation question in order to assure that participants were attentive to the experiment and that they were interpreting the pronoun as coreferential with the relevant noun, as those of the sample in Table 03 below.

Table 03: Stimuli – Experiment 2

antecedent	agreement	context	target	question
<i>Reference Level</i>	<i>Agrees</i>	O <i>indivíduo</i> em questão tinha ficado conhecido como <i>Pedrao</i> . “The individual.MASC at stake had been known as Pedrao.”	<i>Ele</i> foi <i>acusado</i> de um crime. “He was accused of a crime.”	Quem foi acusado de um crime? “Who was accused of a crime?”
<i>Reference Level</i>	<i>Disagrees</i>	O <i>indivíduo</i> em questão tinha ficado conhecido como <i>Pedrao</i> . “The individual.MASC at stake had been known as Pedrao.”	<i>Ela</i> foi <i>acusada</i> de um crime. “She was accused of a crime.”	Quem foi acusada de um crime? “Who was accused of a crime?”
<i>Antecedent Mismatch</i>	<i>Grammatical Agreement</i>	O <i>indivíduo</i> em questão tinha ficado conhecido como <i>Manuela</i> . “The individual.MASC at stake had been known as Manuela.”	<i>Ele</i> foi <i>acusado</i> de um crime. “He was accused of a crime.”	Quem foi acusado de um crime? “Who was accused of a crime?”
<i>Antecedent Mismatch</i>	<i>Contextual Agreement</i>	O <i>indivíduo</i> em questão tinha ficado conhecido como <i>Manuela</i> . “The individual.MASC at stake had been known as Manuela.”	<i>Ela</i> foi <i>acusada</i> de um crime. “She was accused of a crime.”	Quem foi acusada de um crime? “Who was accused of a crime?”

3.2.3. Predictions

Following previous experiments reported in the literature, the empirical predictions with respect to the acceptability judgments were the following: (i) if participants rely on the morphosyntactic features of the antecedent, target sentences grammatically agreeing with the subject of the previous will be judged better than those with pronouns disagreeing with these subjects; (ii) if participants take the semantic and pragmatic features as decisive for the agreement relations between antecedents and pronouns, target sentences agreeing with the contextual gender given by the predicative will be judged better than those which do not agree; and, finally, (iii) if both the morphosyntactic and the semantic-pragmatic features contribute for the acceptability of the sentences, the Reference Level in which the pronoun agrees with the antecedent will be judged the best, the pairs in the *Antecedent Mismatch* conditions will be worse, but better than the *Reference Level* in which the pronoun disagrees with the antecedent in the context sentence.

3.2.4 Results

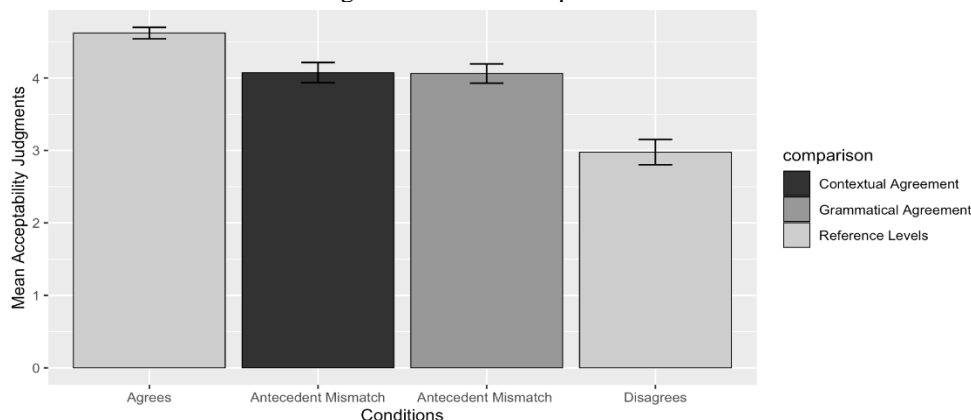
Forty-one (21 female) Brazilian Portuguese speakers living in the southern area of Brazil took part in the experiment. From this group, five participants (three female and two male) were excluded because they reported working in an area related to linguistics or language (translation, literature or linguistics). No participant judged control unacceptable sentences higher than two and a half in average. Trials of participants who read the context and the target sentences faster than 650ms or slower than 9000ms were excluded. More than ninety-eight percent of the answers to the interpretation questions were correct. Trials with incorrect answers were thus excluded from the final data set. The remaining data was then analyzed, as presented in the next sections.

3.2.4.1. Acceptability Judgments

Figure 4 summarizes the results of the experiment. As expected, participants judged the conditions in which the grammatical gender of the common noun and the context gender converged (*Reference Levels*): the sentences were judged better when the pronoun agrees with the gender of the common noun (average: 4.6) and worse when the pronoun disagreed with this gender (average: 2.9). When the genders in the context sentence had a mismatch (*Antecedent Mismatch*), participants graded lower than in the fully convergent sentences, but higher than the

Disagrees Reference Level. In both mismatched conditions (*Contextual* and *Grammatical Agreement*), the sentences agreeing with the morphosyntactic gender averaged 4.

Figure 4: Results – Experiment 2



Acceptability judgments were entered into three log-linear mixed-effects model analyses with random slopes (BARR et al., 2013). Two of them compared all four conditions: in the first one, *Reference Level Agrees* was taken to be the intercept and the other levels were individually compared to it; while in the other one, the intercept of the model was the *Reference Level Disagrees* and the other levels were also individually compared to it (the last row in this model is omitted because it is identical to the last row of the first model). Finally, we fitted an independent model only with *Antecedent Mismatch* conditions (*Contextual Agreement* vs. *Grammatical Agreement*) to have all possible pairwise comparisons. The outcome of the models is summarized in Table 04. Bonferroni corrected coefficient for multiple comparisons are also reported. Contextual and Grammatical Agreement conditions are significantly different from both Reference Levels, but there is no significant difference between the conditions.

Table 04: Log-linear mixed-effects models of Experiment 2 and Bonferroni corrected p-values

Plate 1: M0					Plate 3: M2				
Fixed Effects	Estimate	S.E.	T-val	Pr(> t)	Fixed Effects	Estimate	S.E.	T-val	Pr(> t)
<i>RL/Agrees</i>	4.631	0.067	68.95	< 2e-16	<i>Grammatical Agreement</i>	4.075	0.12	32.0	< 2e-16
<i>Contextual Agreement</i>	-0.544	0.118	-4.58	3.56e-05	<i>Contextual Agreement</i>	0.006	0.15	0.04	0.965
<i>Grammatical Agreement</i>	-0.535	0.143	-3.72	0.00052		1			
<i>RL/Disagrees</i>	-1.648	0.224	-7.34	6.29e-09					

Plate 2: M1					Plate 4: Bonferroni corrected p-values			
Fixed Effects	Estimate	S.E.	T-val	Pr(> t)	Fixed Effects	M0	M1	M2
<i>RL/Disagrees</i>	2.983	0.216	13.80	< 2e-16	<i>Ref. Level</i>	0.000000	0.000000	--
<i>Contextual Agreement</i>	1.104	0.179	6.13	3.59e-07	<i>Contextual Agreement</i>	0.000321	0.000003	0.000000
<i>Grammatical Agreement</i>	1.112	0.208	5.34	2.04e-06	<i>Grammatical Agreement</i>	0.004761	0.000018	1.000000
					<i>Ref. Level</i>	0.000000	--	--

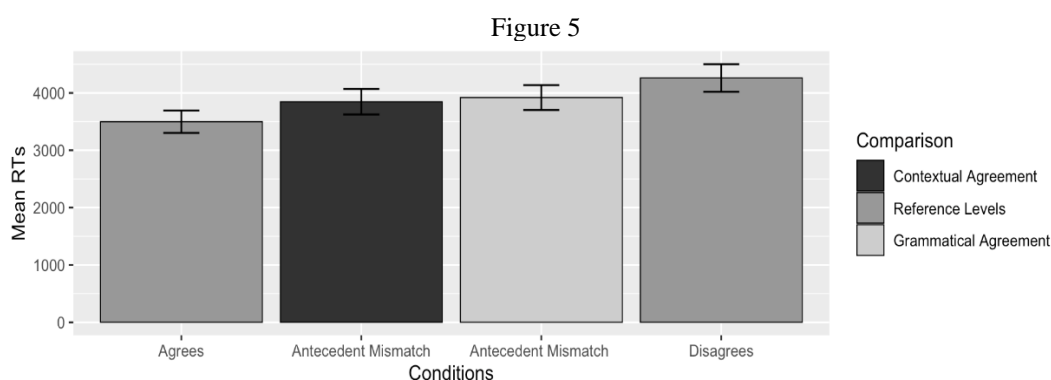
3.2.4.2. Post-hoc Test – Reaction Times of Target Sentences

Given that previous research investigated processing measures (mainly reaction times), we decided to test and report RTs of participants performing the acceptability judgement task⁴. These

⁴ Given the results of the judgment task, we decided to analyze the time spent participants took to pass from the context sentence to the target sentence. We aimed at checking whether some specific lexical items could be easier to process than others, especially when interacting with the mismatching with the semantic-pragmatic gender given by the context.

measures are not thought to answer questions on language processing directly, and we propose to look at them as a first approximation.

Outliers above 10s were also excluded from the sample. Data was averaged by participant and data points beyond three times the standard deviation were also excluded. The results are plotted in Figure 5. We see the inverse pattern of acceptability judgements: in the case where the grammatical and the contextual genders in the context sentence were convergent (*Reference Levels*), participants took more time to read sentences that had a pronoun with a disagreeing gender (average: 4250ms) than when it agreed with the gender of the antecedent (average: 3500ms). When there was a mismatch in the context sentence (*Antecedent Mismatch*), RTs in the target sentences were shorter than the Reference Level with a disagreeing pronoun, but longer than the Reference Level with an agreeing pronoun. When the context sentence had a mismatch (*Antecedent Mismatch*), both conditions (Contextual and Antecedent Mismatch) had similar RTs (averaging 3850ms and 3800ms respectively).



We entered the RTs in a generalized linear mixed-effects models with participants and items as random factors (BARR et al., 2013). As summarized in Table 05 below, the results are very similar to what we observed for the acceptability judgements. The only difference is that Bonferroni corrected p-values show that Contextual Agreement is not significantly different from the Reference Level with the agreeing pronoun, and that Grammatical Agreement is not significantly different from the Reference Level with the disagreeing pronoun.

Table 04: Log-linear mixed-effects models of *post-hoc* test of Experiment 2 and Bonferroni corrected p-values

Plate 1: M0					Plate 3: M2				
Fixed Effects	Estimate	S. E.	T-val	Pr(> t)	Fixed Effects	Estimate	S. E.	T-val	Pr(> t)
Grammatical Agreement	3915.09	228.38	17.143	< 2e-16	Grammatical Agreement	3915.09	228.38	17.143	< 2e-16
Contextual Agreement	-116.91	159.88	-0.731	0.47	Contextual Agreement	-116.91	159.88	-0.731	0.47

Plate 2: M1					Plate 4: Bonferroni corrected p-values			
Fixed Effects	Estimate	S. E.	T-val	Pr(> t)	Fixed Effects	M0	M1	M2
Grammatical Agreement	3915.09	228.38	17.143	< 2e-16	Ref. Level	0.000000	0.000000	--
Contextual Agreement	-116.91	159.88	-0.731	0.47	Contextual Agreement	0.202974	0.035580	0.000000
					Grammatical Agreement	0.008550	0.235680	1.000000
					Ref. Level	0.000852	--	--

The results of these tests are reported in the Appendix. Having found no significant interaction between items and factor antecedent in the RT in the context sentence, we additionally analyzed the RT of the target sentence.

3.2.5. Discussion

As shown above, there is an effect of both contextual and grammatical agreement on acceptability judgments, which seems also to be loosely correlated to RTs. We thus disentangled the effects of contextual and grammatical gender, which contribute individually to the choice of the gender of the pronoun. In convergent contexts (*Reference Levels*), gender violations cause a strong acceptability decline. In mismatch contexts, each violation causes a decline in the acceptability of the sentence, but the speakers still have either gender as an option. This mismatching is correlated with additional difficulties to process the sentences as shown by the increased RTs, especially in condition Grammatical Agreement, in which the contextual gender is not matched.

4. GENERAL DISCUSSION

Crosslinguistically, it has been observed that gender is a grammatical feature without a well-defined semantic counterpart (HOCKETT, 1958; DIXON, 1968; FRANCESCHINA, 2005). As so, assuming the idea that agreement is a “displacement of information” (CORBETT, 2006), it would be expected that only morphosyntactic features would affect the acceptability of agreement relations in general. That is exactly what our Experiment 1 showed: against the Agreement Hierarchy (CORBETT, 1979, 1991, 2006), our results show that, regardless of having a semantic gender (or more generally a specified referent), disagreeing pronouns are less acceptable than agreeing pronouns. Specifically, disagreeing pronouns referring back to nonspecific antecedents (those that denote the class of individuals) are judged as less acceptable than those referring back to specific antecedents (those that denote a specified individual). These results are in line with Alves (2014)’s observation, according to which grammatical gender is a crucial factor in pronoun resolution.

Against Lawall, Maia and Amaral (2012), our results do not depend on feature sharing between the grammatical and the semantic components, since nonspecific antecedents do not necessarily have their semantic gender identical to the grammatical gender. Looking only at the results of Experiment 1, we might conclude that only grammatical agreement is relevant for gender matching between pronouns and their antecedents. However, the literature on Brazilian Portuguese and other gender-marked languages has previously shown that this is not the full picture (CORREA, 2001). Experiment 2 shows results in the opposite direction: both grammatical and contextual gender exert effects on the acceptability and total RT of the target sentences, in which personal pronouns were matched or mismatched with antecedents’ grammatical and/or contextual gender. While corroborating Correa (2001)’s results and some of the results from Alves (2014), results of Experiment 2 go against what was found by Cacciari, Corradini, Padovani & Carreiras (2011) for Italian, according to whom effects of context are diminished when common nouns refer to both biological genders and grammatical agreement predominates.

Differently from what is proposed by Lawall, Maia & Amaral (2012), our proposal is that grammatical features are also computed and equally contribute to the acceptability and processing of coreference between personal pronouns and their antecedents. In fact, both grammatical and contextual features seem to be computed in parallel, and mismatches in each of them, such as agreement violations, are a type of penalty which adds up to lower acceptability and higher difficulty of processing, which is translated into difficulties in integration up to completely non-interpretability. Difficulties in integration, however, do not directly lead speakers to non-interpretability or unacceptability, since violations in only one of the components of agreement is still “good-enough” to get coreferential pronoun interpreted. In cases in which a violation in either grammatical or contextual agreement is inevitable, as in Experiment 2, speakers condescend and cooperate to interpret the sentence.

What is perhaps more surprising than the findings regarding contextual agreement is that grammatical information passes from one sentence to the other independent of semantics. The results reported here are consistent in this regard: Experiments 1 and 2 show that grammatical gender and agreement violations do consistently affect the acceptability of personal pronouns

across sentence boundaries. These results call for a model in which not only contextual and semantic agreement information are stored in working memory, but also grammatical features are stored and are available for convergence checking.

To account for our results, we propose to combine insights from both the Representational Model and from the Retrieval Model: from Representational Models, the parallel features that are, for example, used in HPSG analyses are crucial to understand the cues that are used to compute the agreement and anaphoric relations between a personal pronoun and its antecedent. Notice that, as originally proposed by Pollard and Sag, at least three levels of information seem to be at stake in agreement relations: a grammatical level, a semantic level (such as “index” matching) and a pragmatic level, which includes real-world information, such as stereotypical attributes (CORREA, 2001), stereotypical predicates (ALVES, 2014) and stereotypical proper names in our results. On the other hand, the cumulative and the “good enough” effects may be better understood under a Retrieval Model, in which agreement relations are established “on the fly”, based on the probability of cue-matching. In addition, intersentential agreement and pronoun resolution are inextricably tied to memory because information from a previous sentence needs to be somehow stored and integrated in the current sentence processing. How fast different levels of linguistic information are integrated “on the fly” is a matter of much debate. Correa (2001) and Alves (2014) suggest that pragmatic information is slower than other sorts of information. Our results on processing are total response times, so we are not able to draw clear conclusions on speed of integration. More experiments are needed to understand the differences between levels of linguistic information in online and offline measures.

To conclude, our results have consequences for Representational Models, which suggests that we cannot reduce gender agreement between pronouns and their antecedents to one single level of linguist information. Also, probably we will not be able to establish hard constraints on these agreement relations, since performance and processing factors seem to be decisive in this phenomenon (as also proposed by JOHNSON 2014). Finally, our results also have consequences for Retrieval Models, since we observed that both grammatical and contextual features influence gender agreement between pronouns and their antecedents. Apparently, grammatical and contextual features also have similar weights when both of them are explicitly available in the linguistic context. More research is needed, however, to assure that these features are computed at the same time.

REFERENCES

- AN, A.; ABEILLÉ, A. Agreement and interpretation of binominals in French. In: MÜLLER, S. (ed.). *Proceedings of the 24th International Conference on Head-Driven Phrase Structure Grammar*. Stanford, CA: CSLI Publications, 2017. p. 26-43.
- ALVES, M. C. d. S. *Processamento do traço de gênero na correferência pronominal com antecedentes sobrecomuns e comuns de dois gêneros no português do Brasil*. Dissertação (Mestrado em Linguística) – Universidade Federal do Rio Janeiro, Rio Janeiro, RJ, Brasil. 2014.
- ALVES, M. C. d. S. Gender features in pronoun resolution processing in Brazilian Portuguese. *Proceedings of the Linguistic Society of America*. v. 1, artigo 27, 1-15, 2016.
- ALVES, M. C. d. S. Agreement effects of gender and number in pronominal coreference processing in Brazilian Portuguese. *Revista de Estudos da Linguagem*, v. 25, n. 3, p. 1327-1366, 2017.
- ALVES, M. C. d. S. As diferenças entre gênero gramatical e gênero semântico na recuperação de antecedentes pronominais em português brasileiro. *Diacrítica*. v. 33, n. 2, p. 89-115, 2019.
- BAAYEN, R. H.; DAVIDSON, D. J.; BATES, D. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, v. 59, p. 390-412, 2008.
- BARR, D. J.; LEVY, R.; SCHEEPERS, C.; TILY, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, v. 68, p. 255-278, 2013.
- BREHM, L.; ALDAY, P. M. A decade of mixed models: It’s past time to set your contrasts. *26th Architectures and Mechanisms for Language Processing Conference (AMLap 2020)*. Alemanha: Potsdam, 2020.
- CACCIARI, C.; CARREIRAS, M.; CIONINI, C. B. When Words Have Two Genders: Anaphor Resolution for Italian Functionally Ambiguous Words. *Journal of Memory and Language*, v. 37, n. 4, p 517-532, 1997.
- CACCIARI, C.; PADOVANI, R. Further evidence on gender stereotype priming in language: Semantic facilitation and inhibition on Italian role nouns. *Applied Psycholinguistics*, v. 28, p. 277-293, 2007.
- CACCIARI, C.; CORRADINI, P.; PADOVANI, R.; CARREIRAS, M. Pronoun resolution in Italian: the role of grammatical gender and context. *Journal of Cognitive Psychology*, v. 23, n. 4, p. 416-434, 2011.

- CARREIRAS, M.; GARNHAM, A.; OAKHILL, J. V.; CAIN, K. The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology*, v. 49A, n. 3, p. 639-663, 1997.
- CASADO, A.; PALMA, A.; PAOLIARI, D. The influence of sex information on gender word processing. *Journal of Psycholinguistic Research*, v. 47, n. 3, p. 1-27, 2017.
- CHOMSKY, N. *Lectures on government and binding*. Dordrecht: Foris, 1981.
- CLARK, S. E.; GRONLUND, S. D. Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, v. 3, p. 37-60, 1996.
- COMRIE, B. Polite plurals and predicate agreement. *Language*, v. 51, n. 2: p. 406-418, 1975.
- CORBETT, G. *Predicate Agreement in Russian*. Birmingham: University of Birmingham, 1979.
- CORBETT, G. *Gender*. Cambridge: Cambridge University Press, 1991.
- CORBETT, G. Gender and gender systems. In: ASHER, R. (ed.) *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press, 1994. p. 1347-1353.
- CORBETT, G. *Agreement*. Cambridge: Cambridge University Press, 2006.
- CORRÊA, L. S. Concordância de gênero no processamento de formas pronominais. *Cadernos de Estudos Linguísticos*, v. 40, p. 77-92, 2001.
- CROFT, W. *Radical Construction Grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press, 2001.
- CROFT, W. Agreement as anaphora, anaphora as coreference. In: BAKKER, D; HASPELMATH, M (eds.). *Languages across boundaries: studies in memory of Anna Siewierska*. Berlin: De Gruyter Mouton, 2013. p. 107-129.
- DILLON, B. *Structured access in comprehension*. Tese (Doutorado em Linguística). University of Maryland, College Park, EUA. 2011.
- DILLON, B. Syntactic memory in comprehension of reflexive dependencies: An overview. *Language and Linguistics Compass*, v. 8, p. 171-187, 2014.
- DILLON, B.; MISHLER, A.; SLOGGETT, S.; PHILLIPS, C. Contrasting intrusion profiles for agreement and anaphora. *Journal of Memory and Language*, v. 69, p. 85-103, 2013.
- DIXON, R. Noun Classes. *Lingua*, v. XXI, p. 105-111, 1968.
- DOWTY, D.; JACOBSON, P. Agreement as a semantic phenomenon. In: POWERS, J.; DE JONG, K. (eds.). *Proceedings of the Fifth Eastern States Conference on Linguistics*. Columbus, OH: Ohio State University, 1989. p. 95-108.
- DRUMMOND, A. IbeX Farm. 2013. Disponível em: <http://spellout.net/ibexfarm/>. Acesso em 30 set. 2020.
- FRANCESCHINA, F. *Fossilized Second Language Grammars: The Acquisition of Grammatical Gender*. Amsterdam, Holanda: John Benjamins Publishing Company, 2005.
- HASPELMATH, M. Argument indexing: a conceptual framework for the syntactic status of bound person forms. In: BAKKER, D; HASPELMATH, M (eds.). *Languages across boundaries: studies in memory of Anna Siewierska*. Berlin: De Gruyter Mouton, 2013. p. 197-225.
- HOCKETT, C. *A course in modern linguistics*, Birmingham, Inglaterra: Macmillan, 1958.
- JOHNSON, C. A. *Deconstructing and Reconstructing Semantic Agreement: A Case Study of Multiple Antecedent Agreement in Indo-European*. Tese (Doutorado em Linguística) – The Ohio State University, Columbus, EUA, 2014.
- KATHOL, A. (1999). Agreement and the syntax-morphology interface in HPSG. In: LEVINE, R.; GREEN, G. (eds.). *Studies in Contemporary Phrase Structure Grammar*. Cambridge, Inglaterra, Cambridge University Press: 1999. p. 223-276.
- KATHOL, A.; PRZEPIÓRKOWSKI, A.; TSENG, J. Advanced Topics in Head-Driven Phrase Structure Grammar. In: BORSLEY, R.; BÖRJARS, K. (eds.). *Non-Transformational Syntax*. Oxford, UK/Cambridge, MA: Blackwell Publishing Ltd, 2011. p. 54-111.
- KUZNETSOVA, A.; BROCKHOFF, P. B.; CHRISTENSEN, R. H. B. *ImerTest Package: Tests in Linear Mixed Effects Models*. *Journal of Statistical Software*, v. 82, n. 13, p. 1-26, 2017.
- LAWALL, R.; MAIA, M.; AMARAL, L. Resolução pronominal com antecedentes sobrecomuns e comuns de dois gêneros em Português Brasileiro como lingual maternal, lingual de herança e como segunda lingual. *Revista Linguística*, v. 8, n. 2, p. 134-158, 2005.
- LEITÃO, M.; PEIXOTO, P.; SANTOS, S. Processamento da co-referência intra-sentencial em português brasileiro. *Veredas*, v. 12, n. 2, p. 50-61, 2008.
- LEWIS, R. L.; VASISHTH, S. An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science*, v. 29, p. 375-419, 2005.
- LEWIS, R. L.; VASISHTH, S.; VAN DYKE, J. A. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, v. 10, p. 447-454, 2006.
- MARTIN, A. E., & MCELREE, B. A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, v. 58, p. 879-906, 2008.
- MATTHEWS, P. H. *Syntax*. Cambridge: Cambridge University Press, 1981.
- MCELREE, B. Sentence comprehension is mediated by content addressable memory structures. *Journal of Psycholinguistic Research*, v. 29, p. 111-123, 2000.
- MCELREE, B.; FORAKER, S.; DYER, LISBETH. Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, v. 48, p. 67-91, 2003.
- PARKER, D.; PHILLIPS, C. Reflexive attraction is selective: Evidence from eye-tracking. *Journal of Memory and Language*, v. 94, p. 272-290, 2017.

- PARKER, D.; SHVARTSMAN, M.; VAN DYKE, J. A. The cue-based based retrieval theory of sentence comprehension: New findings and new challenges. In: ESCOBAR, V.; TORRENS, V.; PARODI, T. (eds.). *Language Processing and Disorders*. Cambridge Scholars, Newcastle, Inglaterra, 2017. p. 121-144.
- PATIL, U.; VASISHTH, S.; LEWIS, R. L. Retrieval interference in syntactic processing. *Frontiers of Psychology*, v. 7, artigo 329, 2016.
- POLLARD, C.; SAG, I. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press, 1994.
- SCHAD, D.; HOHENSTEIN, S.; VASISHTH, S.; KLIEGL, R. How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, v. 110, artigo 104038, 2019.
- SLEVC, L. R.; MARTIN, R. C. Syntactic agreement attraction reflects working memory processes. *Journal of Cognitive Psychology*, v. 28, p. 773-790, 2016.
- STEELE, S. The category AUX as a language universal. In: GREENBERG, J.; FERGUSON, C.; MORAVCSIK, E. (eds.). *Universals of human language*. v. 3. Stanford, Stanford University Press: 1978. p. 7-45.
- VAN DYKE, J.; MCELREE, B. Retrieval interference in sentence comprehension. *Journal of Memory and Language*, v. 55, p. 157-166, 2006.
- VAN DYKE, J.; MCELREE, B. Cue-dependent interference in comprehension. *Journal of Memory and Language*, v. 65, p. 247-263, 2011.
- VIGLIOCCO, G.; NICOL, J. Separating hierarchical relations and word order in language production: is proximity concord syntactic or linear? *Cognition*, v. 68, B13-B29, 1998.
- VILLAVICENCIO, A.; SADLER, L.; ARNOLD, D. An HPSG Account of Closest Conjunct Agreement in NP Coordination in Portuguese. In: MÜLLER, S. (ed.). *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar*. Stanford, CA: CSLI Publications, 2005. p. 427-447.
- WECHSLER, S.; ZLATIĆ, L. A Theory of Agreement and its application to Serbo-Croatian. *Language*, v. 76, n. 4, p. 391-423, 2001.

Appendix: Post-hoc Test 2 – Reaction Times of Context Sentences of Experiment 2

Given the results of the judgment task in Experiment 2, we analyzed the time spent by the participants to pass from the context sentence to the target sentence. We aimed at checking whether some specific lexical items could be easier to process than others, especially when interacting with the mismatching with the semantic-pragmatic gender given by the context. In order to do so, we entered the RTs in thirty-six log-linear mixed-effects models with Antecedent (by dummy coding, Convergent 0 or Mismatch 1) and each Item (from 01 to 36) as fixed effects and participants as a random factor. We set the intercept to be an average across all items and factor Antecedent at the level convergent 0 (*i. e. Reference Level*). The output of the models is summarized in Table 05 below (we did not include each item as an individual factor because we are interested only in the interaction with the antecedent condition; the variation across items is expected and is irrelevant for our argumentation in this paper).

Table 05: Linear Mixed-effects Models for RTs

<i>noun</i>	<i>match</i>	<i>mismatch</i>	<i>Fixed effects</i>	<i>Estimate</i>	<i>S. E.</i>	<i>T-val</i>	<i>Pr(> t)</i>
<i>indivíduo(m)</i>	Pedrao	Manuela	ant:item_01	-5.842	511.531	-0.011	0.990891
<i>criança(f)</i>	Aninha	Luisinho	ant:item_02	578.231	655.517	0.882	0.377946
<i>vítima(f)</i>	Elisa	Manuel	ant:item_03	784.978	728.310	1.078	0.281388
<i>testemunha(f)</i>	Paula	Paulo	ant:item_04	778.830	674.926	1.154	0.248806
<i>visita(f)</i>	Magali	Carlos	ant:item_05	555.532	662.094	0.839	0.401651
<i>estrela(f)</i>	Shayane	Ricky	ant:item_06	316.390	659.933	0.479	0.631745
<i>monstro(m)</i>	Josuel	Anita	ant:item_07	191.080	697.025	0.274	0.784039
<i>membro(m)</i>	Leonidas	Gertrudes	ant:item_08	713.469	691.235	1.032	0.302252
<i>ídolo(m)</i>	Joca	Bia	ant:item_09	436.098	662.641	0.658	0.510620
<i>cônjuge(m)</i>	Mário	Josefa	ant:item_10	508.594	710.534	0.716	0.474295
<i>gênio(m)</i>	Rafael	Rafaela	ant:item_11	1100.593	701.533	1.569	0.117011
<i>algoz(m)</i>	Ernesto	Henriqueta	ant:item_12	1168.706	681.465	1.715	0.086668.
<i>carrasco(m)</i>	Roberto	Suelen	ant:item_13	481.273	670.041	0.718	0.472765
<i>criatura(f)</i>	Iara	Ícaro	ant:item_14	392.675	658.343	0.596	0.551010
<i>defunto(m)</i>	Jorge	Cecília	ant:item_15	951.217	704.480	1.350	0.177255
<i>neném(m)</i>	Lucas	Vitória	ant:item_16	695.187	676.290	1.028	0.304235
<i>bebê(m)</i>	José	Clara	ant:item_17	561.977	662.410	0.848	0.396439
<i>ser(m)</i>	Ubiratan	Caiçara	ant:item_18	646.028	679.662	0.951	0.342092
<i>assombração(f)</i>	Kátia	Fausto	ant:item_19	542.687	689.578	0.787	0.431484
<i>pessoa(f)</i>	Joana	Claiton	ant:item_20	58.250	684.623	0.085	0.932212
<i>fantasma(m)</i>	Augusto	Elisabete	ant:item_21	587.337	670.162	0.876	0.381025
<i>tartaruga(f)</i>	Léa	Bob	ant:item_22	84.629	657.647	0.129	0.897634
<i>águia(f)</i>	Nana	Capitão	ant:item_23	1057.778	664.105	1.593	0.111538
<i>baleia(f)</i>	Katrina	Max	ant:item_24	475.771	705.492	0.674	0.500229
<i>girafa(f)</i>	Flor	Tambor	ant:item_25	-3.799	667.186	-0.006	0.995458
<i>mascote(m)</i>	Merlin	Eudora	ant:item_26	-527.561	656.418	-0.804	0.421774
<i>foca(f)</i>	Gisele	Olavo	ant:item_27	-112.557	651.234	-0.173	0.862817
<i>zebra(f)</i>	Leona	Pégasus	ant:item_28	319.071	708.574	0.450	0.652595
<i>polvo(m)</i>	Nemo	Deise	ant:item_29	196.292	691.840	0.284	0.776682
<i>ave(f)</i>	Úrsula	Crístofer	ant:item_30	14.838	657.711	0.023	0.982005
<i>raposa(f)</i>	Bela	Roger	ant:item_31	527.856	658.183	0.802	0.422760
<i>golfinho(m)</i>	Aladin	Jasmine	ant:item_32	595.513	714.442	0.834	0.404748
<i>papagaio(m)</i>	Zeca	Rita	ant:item_33	544.785	667.537	0.816	0.414637
<i>pássaro(m)</i>	Hércules	Serafina	ant:item_34	934.806	638.913	1.463	0.143767
<i>borboleta(f)</i>	Esmeralda	Fígaro	ant:item_35	-323.965	670.701	-0.483	0.629190
<i>onça(f)</i>	Rapunzel	Sebastian	ant:item_36	177.490	684.306	0.259	0.795403

As shown in Table 05, there is only a marginal effect of the interaction between Item 12 and Factor Antecedent. All other items absolutely do not show any significant effect of the interaction between item and type of antecedent on the RT to pass from the first to the second sentence when compared to the average across all items.

Recebido: 12/10/2020

Aceito: 26/8/2021

Publicado: 22/10/2021