frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 14 November 2024
DOI 10.3389/flang.2024.1473948

OPEN ACCESS

EDITED BY
Elena Barbieri,
Northwestern University, United States

REVIEWED BY
Silvia Martínez Ferreiro,
University of A Coruña, Spain
Shinri Ohta,
Kyushu University, Japan
Aya Meltzer-Asscher,
Tel Aviv University, Israel

*CORRESPONDENCE
Paula Luegi
✉ paularibeiro@edu.ulisboa.pt

RECEIVED 31 July 2024
ACCEPTED 25 October 2024
PUBLISHED 14 November 2024

CITATION
Luegi P, Leitão M, Avila-Varela D, Gomes J and Costa A (2024) Reflexive pronoun resolution in Portuguese: testing similarity-based interference. *Front. Lang. Sci.* 3:1473948. doi: 10.3389/flang.2024.1473948

Reflexive pronoun resolution in Portuguese: testing similarity-based interference

Paula Luegi[1]*, Márcio Leitão[2], Daniela Avila-Varela[1,3], Jéssica Gomes[1] and Armanda Costa[1]

[1]Center of Linguistics, School of Arts and Humanities, University of Lisbon, Lisbon, Portugal, [2]Linguistic Processing Laboratory (LaPROL), Federal University of Paraíba, João Pessoa, Brazil, [3]Institute for the Future of Education, Tecnologico de Monterrey, Bilbao, Spain

In the present study, we test whether, during reflexive pronoun resolution, structural cues guide both the language processing system and its underlying memory-based mechanisms or whether the latter might be influenced by non-structural cues as well. Specifically, we explore the inhibitory effects caused by similarity-based interference, which may lead to disruption during reading, reflected in slower reading times and lower accuracy rates. We contrast conditions in which two referents, the reflexive antecedent, and a distractor, are of the same or different gender in sentences with a gender-unmarked reflexive, a gender-marked reinforcement reflexive form, or both. The different types of reflexive constructions allow us to tease apart encoding and retrieval interference since while encoding interference is expected both with gender-marked and gender-unmarked reflexives, retrieval interference is only expected with gender-marked reflexives. In two self-paced reading experiments, one in European Portuguese (EP) and one in Brazilian Portuguese (BP), we find strong and consistent offline results that point toward encoding similarity-based interference. However, the online results only partially support this perspective: In EP, we find encoding interference in the gender-unmarked reflexive and the post-critical regions, while in BP, the effect is only marginally significant in the post-critical region. In addition, in BP, but not in EP, we consistently observe the effects of the participants' accuracy on reading time, with less accurate readers being consistently faster. Overall, our results show that during reflexive pronoun resolution, memory interference can have a negative impact, both during online (reading time) and offline (comprehension accuracy) language processing. With the present study, we contribute to the literature by expanding the set of the tested languages and with more evidence of encoding similarity-based interference, not driven by retrieval cues, on language processing. Moreover, our results are in line with previous studies replicating an asymmetry between robust offline results and elusive online effects. Also, in line with previous studies, our results show that similarity-based interference in grammatical sentences is subtle and may easily be hidden by the large variability between participants (e.g., mean accuracy).

KEYWORDS

reflexive pronoun resolution, Portuguese gender-marked and gender-unmarked reflexives, similarity-based interference, Binding Principle A, encoding and retrieval interference

1 Introduction

Extracting the meaning of a sentence demands the fast and incremental computation of relations between the different words as the sentence unfolds. While, in some cases, a specific cue may be of crucial relevance for the process, the same cue may be irrelevant in others. To illustrate this, we can compare sentences (1) and (2) with sentences (3) and (4).

(1) Mary said that Peter blamed himself.

(2) John said that Peter blamed himself.

(3) Mary saw Peter while he was having a coffee.

(4) John saw Peter while he was having a coffee.

While in (1) and (2) the reflexive pronoun ("himself") can only refer to "Peter," the personal pronoun ("he") can be interpreted as referring to "Peter," in (3) and (4), but also to "John," in (4). Moreover, while gender information is crucial to disambiguate the personal pronoun in (3), leading to ambiguity in (4), it is not useful for the reflexive pronoun even in sentences such as (1) and (2), although in (2) there are two NPs matching the gender of the pronoun, it is unambiguous, as the antecedent is structurally determined. Actually, from a strictly linguistic point of view, proposals such as the Binding Theory (Chomsky, 1981) assume that pronouns are governed by syntactic constraints that restrict what the reflexive might refer to ("Peter," in the presented examples). The question that has been raised and to which there is no definite answer yet is whether, during reflexive pronoun resolution, only syntactic constraints are considered while irrelevant cues such as gender are suppressed or not.

However, the studies that have been conducted to answer this question yielded mixed results: While there are studies that point toward the strict use of syntactic constraints during reflexive pronoun resolution (e.g., Nicol and Swinney, 1989), there are others that show slower reading times in sentences similar to the one presented in (2), in which there are two nouns with the same gender (e.g., Badecker and Straub, 2002). As we will explain in the next section, the effects detected in sentences such as (2) are explained by interference from memory-based mechanisms that might have an impact on language processing.

In the present study, we aim to examine whether reflexive pronoun resolution might be impacted by memory limitations, such as similarity-based interference, or if, instead, it is solely guided by syntactic constraints, being immune to memory interference. In the following sections, we will illustrate how memory is expected to interfere with language processing and describe the cue-based retrieval model proposed by Lewis and Vasishth (2005), which accounts for memory interference during language processing. We will focus then on studies that have analyzed the impact of memory interference during pronoun resolution. Before explaining in detail the study we conducted, we will briefly present the relevant properties of the Portuguese reflexive pronouns, focusing on the two varieties tested in the present study: European and Brazilian Portuguese (henceforth, EP and BP).

1.1 Similarity-based interference in language processing

Although language comprehension is usually a smooth and successful process, some paradigmatic examples illustrate how it can sometimes be deluded. One such case is *agreement attraction*, one of the different examples of the well-known phenomena known as *grammatical illusion* (e.g., Phillips et al., 2011). For instance, sentences such as "*The key to the cabinets are on the table*," in which the verb ("are"), against the syntactic constraints imposed by agreement, does not agree with the subject ("the key"), are not only frequent in language production but also highly acceptable in language comprehension (for a review, Phillips et al., 2011; Laurinavichyute, 2021).

This "failure" is usually explained as a limitation imposed by memory overload. Since linguistic information needs to be stored, maintained, and retrieved rapidly and incrementally, memory is essential for language comprehension (Lewis et al., 2006). So, although the language system might be guided by linguistic-specific constraints, it also needs to rely on general memory-based mechanisms, being, therefore, prone to memory failure. In the case of the sentence above, the verb will search for a plural subject-noun, and since the plural intervening noun ("the cabinets") partially fills this criterion, it legitimates the agreement-attraction *illusion*.

One model that accounts for the limitations imposed by memory to the language processing mechanism was proposed by Lewis and Vasishth (2005). The model assumes a content-addressable architecture of the memory system, in which items are accessed in memory through feature matching based on their distinguishable features. In this feature-based mechanism, linguistic features, like gender, number, or agreement, are used to code, maintain, and recover linguistic material from memory. Memory is, therefore, structured into different phases: encoding, when linguistic material is coded to be stored in memory; maintenance, which is the process of keeping in memory the perceived material; and retrieval, the process of recovering stored material from memory.

As explained in Lewis et al. (2006), during comprehension, items are stored in memory and represented as a bundle of features. Importantly, only one item at a time is available for processing, and, therefore, as soon as one item is available for processing, the level of activation of the previous item decays (Laurinavichyute, 2021). That means that whenever a dependency between two items needs to be established, the item that needs to be related to the ongoing processing item needs to be reactivated (or retrieved). The retrieval process starts with the set of retrieval features that are activated by the item that triggered the retrieval process, such as, for instance, agreement (number, person) or structural cues (syntactic function: subject, object). Consequently, the set of features available for the retrieval process is just a subset of features of the item to be retrieved since it is composed only of the features activated by the item that triggered the retrieval process. For instance, in a subject-verb dependency, the verb might activate number features only, even though the to-be-retrieved item has both gender and number features in its feature bundles. Moreover, each retrieval cue is blindly spread among all the items that match that specific cue. Broadly speaking, the item that has more matching cues

reaches the highest activation level and is, therefore, selected as the retrieved item.

However, since all the items are encoded, maintained, and retrieved from the memory based on their features, some interference is expected to occur whenever there is some feature similarity, i.e., the so-called similarity-based interference (Oberauer and Kliegl, 2006; Gordon et al., 2002). The Lewis and Vasishth (2005) model predicts that if more than one item matches the retrieval cues, the cue spreading activation is divided equally among all the matching items, and, therefore, all the items with matching features get some degree of activation. This competition might lead to cue overload and, consequentially, to activation decrease. Processing speed and accuracy are, therefore, expected to decrease as the amount of interference increases.

Moreover, although the Lewis and Vasishth (2005)'s model is focused on the retrieval process, as pointed out by Lewis et al. (2006, p. 449), some similarity-based interference might also be expected during sentence processing. As proposed by Oberauer and Kliegl (2006), similarity-based interference can impact both the encoding and retrieval processes, but in distinct ways. During encoding, similarity-based interference is predicted to occur whenever items share similar features, regardless of whether these features are used for retrieval or not. In contrast, retrieval similarity-based interference is expected to happen only when there is a similarity between the features that are needed explicitly for items' retrieval (McElree et al., 2003; Gordon et al., 2002). However, as Villata et al. (2018) point out, although *encoding* and *retrieval* are easily distinguished at a theoretical level, these concepts are complex and difficult to tease apart empirically because the effects of both encoding and retrieval are expected at the moment of retrieval. Due to feature-overwriting, the memory representation of items with shared features is degraded. That is, those items are less distinguishable due to encoding similarity-based interference. Therefore, the base level activation of those items will be lower at the moment of retrieval, leading to a slowdown in processing times (Jäger et al., 2015), not caused by retrieval interference but by encoding interference. To distinguish the two processes it is needed, therefore, to tease them apart by separating cues used only for encoding from cues used for retrieval.

The distinction between the *locus* of the interference effect, tested by Villata et al. (2018), is of crucial interest for our study. Considering the difficulty of identifying encoding or retrieval interference effects in isolation, the authors conducted a study to tease these two aspects apart. Villata et al. (2018) contrasted retrieval and encoding interference by looking at how gender [example (5)], in Italian, or number-matching [examples (6) and (7)], in English, subject and object nouns, interfere with the processing of verbs in object relative clauses in two self-paced reading experiments.

(5) The dancer$_{(masc)}$ that the waiter$_{(masc)}$/waiter$_{(fem)}$ has surprised$_{(\emptyset)}$ drank a cocktail with alcohol.

(6) The dancer$_{(sing)}$/dancers$_{(plural)}$ that the waiter$_{(sing)}$ strongly criticizes$_{(sing)}$ most of the time ordered a rum cocktail.

(7) The dancer$_{(sing)}$/dancers$_{(plural)}$ that the waiter$_{(sing)}$ strongly criticized$_{(\emptyset)}$ most of the time ordered a rum cocktail.

Considering that, in the tested sentences, gender is not a retrieval cue in Italian [example (5)] as the number is not in English in the past tense [example (7)], only effects of encoding can explain any difference between overlap (e.g., masculine–masculine; singular–singular) and no-overlap conditions (e.g., masculine–feminine; plural–singular). On the other hand, retrieval interference is expected to occur in the present tense sentences [example (6)] in the overlap conditions (singular–singular). Villata et al. (2018) found slower reading times and more errors in overlapping conditions, even when there were no retrieval cues, showing, therefore, interference effects of both encoding and retrieval. Off-line results (decrease in accuracy in overlapping conditions) were found in Italian and English, independently of the existence of a retrieval cue, which, in Villata et al. (2018) perspective, reveals that the *locus* of interference is encoding. Online results are, however, weaker than the offline ones and are clearer when there are retrieval cues (English number features in the present tense which the authors assume reflect a constraint imposed by the properties of the language).

The debate about similarity-based interference effects during reflexive pronoun resolution, and also about the distinction between encoding and retrieval interference effects, has gained interest mainly since the study of Dillon et al. (2013). The following sections focus on studies specifically addressing this question.

1.2 Similarity-based interference in reflexive pronoun resolution

Reflexives such as "himself" or "herself" (an anaphora in *stricto sensu*) constitute a particular category of anaphoric expressions. According to syntactic proposals, these expressions are governed by binding principles that stipulate that they can only be co-indexed (co-indexation is represented in the examples with subscript letters: The same letter (or index) identifies an equal reference) with an antecedent within its binding domain (Chomsky, 1981), roughly the local clause. Therefore, accordingly to the Binding Principle A, the reflexive "himself," in (8) and (9), must be bound by, or refer to, "John."

(8) John$_i$ described himself$_i$ as a good person.

(9) Bill$_i$ said that John$_k$ described himself$_{*i/k}$ as a good person.

The syntactic constraints described by Chomsky (1981) have been empirically tested to analyze their direct impact on language processing. The goal is to answer the following question: When processing a reflexive pronoun, does Binding Principle A immediately filter out ungrammatical referents, or are all referents initially considered regardless of their grammatical availability? For instance, in example (9), if the syntactic principles exclude ungrammatical referents, "Bill" will not be available as an antecedent for the reflexive pronoun "himself;" on the other hand, if the syntactic constraints are not strictly followed, both "John" and "Bill" will be considered as antecedents for the reflexive pronoun.

Broadly, there are two opposite approaches: On the one hand, it is assumed that language processing strictly follows the syntactic rules, and, therefore, syntax is considered a central and fundamental aspect of the language processing mechanism; on the

other hand, syntactic rules are one of the many linguistic sources of information that are used during language processing. The former assumes a structural or syntactic-based mechanism of language processing, and we refer to it as the "structure-based account." The second approach expects an interactive use of different linguistic information sources (syntactic rules, morphosyntactic information, or semantic roles), even though these cues may lead to ungrammatical representations. This proposal is referred to here as the "cue-based account." The two perspectives have been tested with reflexives, mainly based on the seminal works of Nicol and Swinney (1989) and Badecker and Straub (2002).

Nicol and Swinney (1989) tested sentences like (10) through different experiments, using probe-word cross-modal priming experiments. Participants had to judge whether a word presented at some point in a sentence was an English word or not. In the example (10), after the reflexive "himself," words related to "the landlord," "the janitor," or "the fireman" could be presented (in addition to unrelated words).

(10) The landlord told the janitor that the fireman with the gas-mask would protect himself.

Shorter reaction times, reflecting a facilitation effect, were found in the probe-word semantically related to the antecedent of the reflexive ("smoke"–"the fireman") when compared to the reaction times to the unrelated words ("shift"–"the fireman") or to words related to the distractors ("rent"–"the landlord," "clean"–"the janitor"). Nicol and Swinney (1989) considered that these results show that only the grammatical antecedent is reactivated immediately after the reflexive. Based on their results, the authors proposed the *initial-filter hypothesis*, a structure-based approach: The syntactic rules filter out the ungrammatical referents, and only the grammatical antecedent for the reflexive remains available for reflexive resolution.

Afterward, Badecker and Straub (2002) conducted a similar study to contrast the *initial-filter hypothesis* with the *interactive parallel-constraint hypothesis*, a cue-based approach. Using the self-paced reading paradigm, the authors tested sentences like (11).

(11) Jane/John thought that Bill owed himself another opportunity to solve the problem.

The results indicate a gender interference effect, with readers taking longer to read conditions where the distractor and the antecedent shared the same gender (e.g., "John" and "Bill") compared to conditions where they were of different gender (e.g., "Jane" and "Bill"). Badecker and Straub (2002) interpreted these results as suggesting that gender information is also used to search for the reflexives' antecedent, and therefore, even ungrammatical antecedents are being considered during reflexive pronoun resolution. Taking into account these results, they proposed that there is evidence for a cue-based approach to language processing.

After these two studies, other studies were conducted to analyze the resolution of reflexive pronouns. The main focus was consistently on contrasting the strict use of syntactic rules with the use of various sources of linguistic information, as well as identifying where the effect occurs within the sentence (for a review, see Jäger et al., 2017). However, the studies yielded mixed results. While some studies found no effects at all (Sturt, 2003; Dillon et al., 2013, Exp. 2), others found an increase in reading time (Badecker and Straub, 2002; Clackson and Heyer, 2014, Exp. 3, 4) in conditions in which the antecedent and the distractor share the same gender ["John"–"Bill," in (11)], and yet others found, also with an antecedent and a distractor of the same gender, faster reading times for the reflexive (Sturt, 2003; Cunnings and Felser, 2013, Exp. 1). Moreover, while in some studies, effects were only found in some words after the reflexive (Badecker and Straub, 2002) or during the reflexive rereading (Sturt, 2003), in others, effects were immediately detected (Clackson and Heyer, 2014). Considering this diversity of results, there was a perspective shift on the focus of analysis of reflexive pronoun resolution mainly driven, as mentioned previously, by Dillon et al. (2013)'s study.

Dillon et al. (2013) contrasted two different types of linguistic dependencies: subject-verb agreement and reflexive anaphoras in English. The goal was to test the relationship between linguistic representation and memory access, analyzing memory-based interference. The authors tested agreement conditions with, as in (12), or without intrusion, as in (13), and reflexive conditions with, as in (14), or without intrusion, as in (15).

(12) The new executive who oversaw the middle managers apparently was dishonest about the company's profits.

(13) The new executive who oversaw the middle manager apparently was dishonest about the company's profits.

(14) The new executive who oversaw the middle managers apparently doubted himself on most major decisions.

(15) The new executive who oversaw the middle manager apparently doubted himself on most major decisions.

While the authors were unable to find any interference effects of the distractor ("manager(s)") in conditions with reflexives, they found interference effects reflected in slower reading times in conditions with subject-verb agreement. The authors concluded that morphological agreement constraints (such as number, in the tested examples) are not uniformly used as retrieval cues across different types of linguistic dependencies. Moreover, they propose that reflexive pronoun resolution is solely guided by syntactic constraints, in contrast to subject-verb agreement dependencies comprehension guided by both morphological and syntactic cues. Therefore, Dillon et al. (2013) propose that the impact of memory interference in online sentence processing depends on the type of linguistic dependency at stake. While some dependencies are prone to memory interference, such as subject-verb agreement, others are solely guided by syntactic constraints, such as reflexive dependencies (Dillon, 2014).

Considering that the lack of results in reflexive pronoun conditions presented in Dillon et al. (2013) could be explained by a lack of statistical power, Jäger et al. (2020) conducted a large-sample replication of Dillon et al. (2013)'s study. In addition to the reanalysis of the original data, with Bayesian models, and of a computational model simulation, the authors conducted a large-sample eyetracking study with the same experimental items as the ones tested by Dillon et al. (2013). Overall, and despite the large number of participants, the results were inconclusive: no

clear inhibitory interference effects were found in the reflexive conditions.[1]

However, most of the studies on pronoun resolution have analyzed only English singular reflexive pronouns. The studies of Jäger et al. (2015) and Laurinavichyute et al. (2017), presented below, are an exception. In these two studies, both gender-marked and gender-unmarked reflexives were tested. This is particularly relevant because, in sentences with two nouns that have gender overlap, while gender-marked reflexives, such as the ones tested in English, can trigger both retrieval and encoding interference, gender-unmarked reflexives can only trigger encoding interference. If gender is not a retrieval cue to search for the reflexive antecedent, then no retrieval interference is expected in gender-overlapping conditions. Therefore, any interference can only be explained by encoding interference.

Jäger et al. (2015) tested the resolution of German reflexives (Experiments 1 and 2), which are gender-unmarked, and of Swedish gender-marked pronominal possessives and gender-unmarked reflexive possessives (Experiment 3). The idea was to contrast encoding and retrieval interference, considering that if reflexives are gender unmarked, gender cannot be used as a retrieval cue. Therefore, any effect (longer reading times and more errors when answering final questions) found in the resolution of the reflexive in conditions with distractor and antecedent of the same gender can only be attributed to encoding interference. Jäger et al. (2015) found no online effects during reflexive pronoun resolution and a facilitation effect in sentences with gender-marked pronominal possessives in Swedish when they matched in gender with the distractor. The authors concluded that in the absence of clear effects of encoding, and also with the evidence of retrieval interference during pronominal resolution, the results (both from their study and previous studies) can only be explained by retrieval interference effects.

Following Jäger et al. (2015), Laurinavichyute et al. (2017) conducted a study with German gender-unmarked reflexives (similar to Exp. 1 of Jäger et al., 2015) and Russian gender-marked and unmarked reflexives. The experiments with Russian (Exp. 2A and 2B) are of particular interest to our study since the structures are similar to the ones we tested. In (16) and (17), we reproduce the glosa-translation and the full translation examples from Laurinavichyute et al. (2017).

(16) Swindler$_{(fem)}$ whom merchant$_{(fem)}$ hires for robbery, self$_{ACC(\emptyset)}$/herself$_{ACC(fem)}$ significantly overestimates in ability to do trickery.

(17) Swindler$_{(fem)}$ whom merchant$_{(masc)}$ hires for robbery, self$_{ACC(\emptyset)}$/herself$_{ACC(fem)}$ significantly overestimates in ability to do trickery.

"The swindler$_{(fem)}$, whom a merchant$_{(fem)/(masc)}$ hires for a robbery, significantly overestimates her own$_{ACC(\emptyset)/(fem)}$ trickery skills."

While the authors replicated the lack of online results for German, they found online and offline effects for Russian in conditions where the antecedent and the distractor shared the same gender [e.g., (16)], with longer reading times and more errors in the reflexive gender-unmarked condition, but no effects when testing the gender-marked reflexives. The authors concluded that the results evidence encoding interference during reflexive pronoun resolution and are inconsistent with retrieval interference.

In the present study, we tested EP and BP reflexive pronouns. Therefore, in the next section, we will briefly present the relevant properties of the reflexive pronouns in these two varieties of Portuguese.

1.3 Reflexives in European and Brazilian Portuguese

In EP and BP, reflexivity is marked with a clitic pronoun that shares person and number features with its antecedent, like in (18), except for the third person pronoun. The third person reflexive pronoun *se* is not number marked, as in (19) and (20), and, as with all the other forms, it is also not gender-marked, as illustrated in (21).

(18) *Eu lavei-me.*
"I washed myself."

(19) *Eles lavaram-se.*
"They washed themselves."

(20) *O João lavou-se.*
"John washed himself."

(21) *A Maria lavou-se.*
"Mary washed herself."

Moreover, reflexive pronoun gender-marked reinforcement forms can also be used, as exemplified in (22). This form is always gender marked with the gender-marked suffix *-a*, feminine, as in (23), or *-o*, masculine, as in (22), at the end of the form *mesm-* (for a more detailed description about EP and BP characteristics on binding in Portuguese, see Menuzzi and Lobo, 2016). The reinforcement form always appears after the verb, within a prepositional phrase, either when the reflexive is in enclisis, as in (22), or in proclisis, as in (24).

(22) *O João*$_{(masc)}$ *lavou-se a si mesmo*$_{(masc)}$.
John washed-SELF himself.
"John washed himself."

(23) *A Maria*$_{(fem)}$ *lavou-se a si mesma*$_{(fem)}$.
Mary washed-SELF herself.
"Mary washed herself."

(24) *O João*$_{(masc)}$/ *A Maria*$_{(fem)}$ *não se lavou*
John/Mary not SELF washed
a si mesmo$_{(masc)}$/ *a si mesma*$_{(fem)}$.
himself.

1 Nonetheless, Jäger et al., 2020 found consistent facilitatory effects in the ungrammatical conditions (*"Mary said that Peter washed herself."). The facilitatory effect is predicted by the cue-based retrieval model and is reflected in faster reading times in the presence of a partially matching distractor. As mentioned by Jäger et al. (2020), while facilitatory effects are persistently found, inhibitory effects in reflexive pronoun resolution are not consistent across studies.

"John/Mary did not wash himself/herself."

There are, however, some differences between EP and BP concerning sentence structure with reflexive clitics. Considering the reflexive position, EP favors enclisis, like in (25), while BP mainly uses proclisis, like in (26).

(25) *O João vestiu-se.*

(26) *João se vestiu.*
"John dressed himself."

Secondly, both EP and BP allow the reinforcement reflexive form. Still, there are differences between the two varieties: While the reinforcement reflexive form obligatorily accompanies the reflexive pronouns in EP, like in (27), in BP, it can be used both with the reflexive, like in (28), or not, as in (29), with verbs that are not inherently reflexive[2] (Menuzzi and Lobo, 2016; Mello, 2009; Pereira, 2007).

(27) *O João vestiu-se/ vestiu*(-se)* *a si mesmo.*
John dressed-SELF/dressed*(-SELF) himself.
"John dressed himself."

(28) *O João se vestiu a si mesmo.*
John SELF dressed himself.
"John dressed himself."

(29) *O João Ø vestiu a si mesmo.*
John Ø dressed himself.
"John dressed himself."

In sum, in Portuguese, it is possible to have sentences containing only gender-unmarked reflexives, sentences containing both gender-unmarked and gender-marked reflexives, and sentences containing only gender-marked reflexives. Therefore, Portuguese is similar to other languages, such as Russian, allowing the use of both gender-unmarked (*se*) and gender-marked (*a si mesmo*) reflexives. However, in contrast to what happens, for instance, in Russian, in Portuguese, the gender-marked reinforcement reflexive form must always appear after the verb, mandatorily preceded by the gender-unmarked reflexive in EP, or not, in BP.

However, it is important to point out that the frequency of these structures is not equivalent within and across each Portuguese variety. As described in Leitão et al. (2017), the default and more frequent forms are the gender-unmarked reflexive pronouns without the reinforcement forms, both in EP and BP. Nonetheless, the authors show that while in BP, when the reinforcement form *a si mesm*(*o*/*a*) is present, the clitic *se* is absent in 74% of the cases, in EP, the omission of the clitic is very infrequent, showing that the use of the form *a si mesm*(*o*/*a*) actually works as a reinforcement phrase in EP.

2 Verbs that are inherently reflexive are *comportar* ("to behave"), *arrepender* ("to regret"), *queixar* ("to blame"). Although these verbs allow the clitic *se* they are not real reflexives. With these verbs, it is not possible to have *a si mesmo/a* as a reinforcement form and, therefore, it is not possible to omit the clitic *se* in BP (see Fonseca, 2012, for further discussion about this topic).

TABLE 1 Examples of reflexive type constructions tested on our study ("John said that Bruno cut himself with the knife").

Unmarked	O João disse que o Bruno […] **se** cortou com o canivete […].	EP & BP
Marked	O João disse que o Bruno […] cortou **a si mesmo** com o canivete […].	BP
Unmarked + Marked	O João disse que o Bruno […] **se** cortou **a si mesmo** com o canivete […].	EP

Short versions of the sentences. No examples of different gender conditions were included.

2 The current study

In the present study, we analyze reflexive pronoun resolution in EP and BP, considering the impact of structural constraints and memory-based mechanisms. We explore the inhibitory effects caused by encoding or retrieval similarity-based interference during language processing, which might lead to disruption during reading. We contrast conditions in which two referents, the antecedent of the reflexive and a distractor, are of the same or different gender in sentences with a gender-unmarked reflexive, a gender-marked reinforcement reflexive form, or both. In Table 1, we present the three types of reflexive constructions we tested in our study (the examples provided are short illustrations in which we omit the phrases between the second noun and the verb and the final phrase; in addition, only examples of sentences with same gender referents are included). A complete list of examples is provided at the beginning of each experiment [(30) to (35)].

The predictions are as follows:

1. The lack of differences between conditions, with same or different gender, that is, no differences or null effects, could be explained as the strict adherence to the syntactic constraints.
2. Differences between conditions with referents of the same gender and conditions with referents of different gender, with gender-unmarked reflexive pronouns, can only be explained by similarity-based encoding interference. Assuming that predictions of similarity-based interference are due to encoding (feature overwrite), inhibitory effects are expected when the two referents share the same gender, leading to slower reading times and lower accuracy rates in these conditions.
3. Differences between conditions with referents of the same gender and conditions with referents of different gender, in the gender-marked reflexive conditions, can be explained by both encoding and retrieval similarity-based interference. Under the predictions of the cue-based retrieval model, inhibitory effects are expected when the two referents share the same gender, leading to slower reading times and lower accuracy rates in these conditions.

In sum, with the design of our experiments, we foresee four different possible scenarios. It is important to highlight that some of these scenarios result from the combination of effects found or not found in different conditions. In all the scenarios in which we assume the existence of effects described below, we expect slower reading times and lower accuracy rates in the conditions with gender overlap between the two presented nouns.

In the first scenario, no effects are found, independently of the tested conditions, and, therefore, we can assume that the syntactic constraints are strictly followed (in line with the structure-based perspective).

The second possibility is to find effects in the gender-unmarked reflexive. The effect of gender, if clearly present when the reflexive is gender-unmarked, can only be explained by similarity-based encoding interference. This scenario would support the existence of similarity-based encoding interference during language processing. The antecedent is harder to retrieve because its' shared encoding features lead to its lower activation level and, therefore, to a slower retrieval process.

The third scenario is to find interference effects in the gender-marked reinforcement conditions, but not in the gender-unmarked ones. This result can only be explained by similarity-based retrieval interference during the search for the reflexive pronoun antecedent. It is important to highlight that this scenario is assumed if no effects of encoding are found in the gender-unmarked conditions. Although the gender-marked reflexives might lead to both encoding and retrieval similarity-based interference, if effects are only found in the gender-marked conditions but not in the gender-unmarked ones, then encoding interference cannot be the explanation for the effects found in gender-marked reflexive forms.

The only way to isolate retrieval interference with the current design would be to show more interference when retrieval and encoding interference are both possible, than when only encoding interference is possible. The difference would be the effect of retrieval.

Finally, the last scenario is to find effects with both types of reflexives. This would reveal similarity-based interference during reflexive pronoun resolution due to encoding or both encoding and retrieval. Moreover, as noted by one reviewer, if effects in the gender-marked reflexive are simultaneously caused by encoding and retrieval, then the effects in this reflexive are expected to be stronger than the effects in the gender-unmarked reflexive, which have only one source of interference (encoding).

In addition to the diverse scenarios presented above, as stressed by one of the reviewers, the structures created allow us to test the retrieval of the antecedent of the reflexive at different regions of the sentences and, more importantly, to test if the retrieval of the antecedent happens every time a reflexive is encountered or re-encountered, as in double reflexive constructions.

Assuming the predictions of the Lewis and Vasishth (2005) model, since any word is allowed to initiate the retrieval process, as long as a linguistic dependency needs to be established, the retrieval of the antecedent is expected to occur on both gender-unmarked and gender-marked reflexives. Therefore, the structures of our study allow us to test different possibilities, particularly with the gender-marked reinforcement forms. Considering that items increase their base-level activation each time they are retrieved, the antecedent of the gender-marked reinforcement form is expected to have a higher activation level when retrieved at the reinforcement form in EP than in BP. Therefore, similarity-based interference effects in the reinforcement form are expected to be lower or even null in the reinforcement form in EP, since the antecedent was already retrieved once (immediately before) or might even not be retrieved again. On the other hand, in BP, every sentence has only one reflexive form, and thus both the gender-unmarked and the gender-marked reflexives will trigger retrieval of the reflexive's antecedent for the first time. We get back to this issue in the General Discussion.

2.1 General methods

This section describes the general aspects of the two experiments presented in the paper. Specific aspects, such as the number of participants and the results, are then presented while describing each experiment.

2.1.1 Materials

The experimental items were created by crossing two factors with two levels each in a 2 × 2 design: Referents' Gender (No-Overlap vs. Overlap) and Reflexive Type (ReflexReinf vs. ReflexOnly, in EP, and ReinfOnly vs. ReflexOnly, in BP). In the Referents' Gender we manipulated the gender of the referents introduced in the sentence, and in the Reflexive Type we manipulated the type of reflexive presented in the sentence.

In the Referents' Gender condition, we included sentences in which the gender of the antecedent and the distractor either overlap or do not overlap. Therefore, there are two conditions: Overlap (30) and No-Overlap (31).

(30) *O João$_{(masc)}$ garantiu que o Bruno$_{(masc)}$ durante a visita se cortou com o canivete do jardim.*

(31) *A Ana$_{(fem)}$ garantiu que o Bruno$_{(masc)}$ durante a visita se cortou com o canivete do jardim.*
"John$_{(masc)}$ / Anne$_{(fem)}$ assured that Bruno$_{(masc)}$ during the visit SELF cut with the knife of the garden."

Within the Reflexive Type condition, two types of sentences were included, which differed in EP and BP. In EP, half of the sentences exclusively contained the gender-unmarked reflexive pronoun in proclisis *se VERB* [as exemplified in (32)], referred to here to as the ReflexOnly condition. The remaining half of the sentences had the gender-unmarked reflexive pronoun in proclisis and the gender-marked reflexive reinforcement form in enclisis, specifically *se VERB a si mesmo/a* [as exemplified in (33)], the ReflexReinf condition.

(32) *O João$_{(masc)}$ / A Ana$_{(fem)}$ garantiu que o Bruno$_{(masc)}$ durante a visita se cortou com o canivete do jardim.*

(33) *O João$_{(masc)}$ / A Ana$_{(fem)}$ garantiu que o Bruno$_{(masc)}$ durante a visita se cortou a si mesmo$_{(masc)}$ com o canivete do jardim.*
"John$_{(masc)}$ / Anne$_{(fem)}$ assured that Bruno$_{(masc)}$ during the visit SELF cut (himself$_{(masc)}$) with the knife of the garden."

In BP, half of the sentences exclusively contained the gender-unmarked reflexive pronoun in proclisis *se VERB*, the ReflexOnly condition [as exemplified in (34)]. The remaining half of the sentences exclusively contained the gender-marked reflexive reinforcement form in enclisis, specifically *VERB a si mesmo/a* [as exemplified in (35)], the ReinfOnly condition.

(34) *O João$_{(masc)}$ / A Ana$_{(fem)}$ garantiu que o Bruno$_{(masc)}$ durante a visita se cortou com o canivete do jardim.*

(35) *O João$_{(masc)}$ / A Ana$_{(fem)}$ garantiu que o Bruno$_{(masc)}$ durante a visita cortou a si mesmo com o canivete do jardim.*
"John$_{(masc)}$ / Anne$_{(fem)}$ assured that Bruno$_{(masc)}$ during the visit (SELF) cut (himself$_{(masc)}$) with the knife of the garden."

The experimental items were complex sentences composed of a main clause with an embedded complement clause (always a verb complement). In the main clause, the subject always served as the distractor [*João*, in (30)], while, within the complement clause, the subject functioned as the antecedent, i.e., the only grammatical referent [*Bruno*, in (30)] of the object reflexive pronoun.

Although we initially considered using sentences similar to the ones tested in other studies, such as the one used in Laurinavichyute et al. (2017), for instance, considering the design and especially to keep the sentences as similar as possible in EP and BP, we opted, instead, for complement clauses. This decision was mainly motivated by the following reason: (i) both the distractor and the antecedent have the same syntactic function, leading to higher feature overlap (e.g., gender and syntactic function, in overlap conditions); (ii) it allows a direct comparison between EP and BP, namely in complement clauses, the reflexive always precedes the verb in both varieties, unlike other syntactic structures (e.g., "The young man that the teacher saw yesterday cut himself"; in EP: *O jovem que o professor viu ontem cortou-se*, in BP: *O jovem que o professor viu ontem se cortou*); and (iii) this configuration cancels the reactivation of any referent before reaching the anaphora, which could otherwise impact the results. According to Laurinavichyute et al. (2017), when the verb is encountered, the subject is activated, which could create noise in the processing of the reflexive if it were not in proclisis.

We used proper nouns for the referents since, in Portuguese, proper nouns possess clear gender distinctions, either feminine or masculine, and are preceded by a definite article marked for gender. We counterbalanced the gender of the distractor and the antecedent in the sentences within the Overlap condition. This resulted in presenting half of the Overlap condition sentences with two feminine nouns for the distractor and antecedent (e.g., *Ana* and *Sofia*), while the other half featured two masculine nouns (e.g., *João* and *Bruno*). Similarly, in the No-Overlap condition, half of the sentences included a feminine noun for the distractor (e.g., *Ana*) and a masculine noun for the antecedent (e.g., *Bruno*), and the other half had the reversed order (masculine distractor and feminine antecedent).

In addition, a phrase such as *durante a visita* ("during the visit") was inserted between the antecedent and the reflexive to minimize recency effects and avoid the antecedent being presented immediately before the reflexive.

The experimental items were selected based on an acceptability judgment pre-test. The pre-test was conducted with a separate sample of participants to verify the overall acceptability of the sentences. The experiment was constructed and presented in OnEXP.[3] Only the EP items were tested, and, therefore, all the participants of the pre-test were native speakers of EP. Fifty items were tested in the two Reflexive Type conditions: with only gender-unmarked reflexive and with both the gender-unmarked reflexive and the gender-marked reinforcement form. Each sentence was evaluated on a 5-point Likert Scale (where 1 was completely unacceptable and 5 was completely acceptable). Participants were asked to comment on their classification if their rate was 1 or 2, but this was not mandatory. Comments allowed us to understand what could be less acceptable in the sentence, allowing for adjustments if necessary.

Overall, sentences with the gender-marked reinforcement reflexive form were judged less acceptable (*M*: 2.50, range: 1.37–3.53) than sentences with the gender-unmarked reflexive (*M*: 3.65, range: 2.69–4.38). The reinforcement form has an emphatic purpose, which might explain its low acceptability rate in the written mode. From the 50 items tested in the pre-test, we selected 32 items with the highest mean acceptance rate: $M = 3.21$ [$M = 3.62$ (2.70–4.38) for gender-unmarked reflexives, and $M = 2.79$ (1.92–3.53) for gender-marked reinforcement sentences].

Although the pre-test was only conducted in EP and with EP speakers, all the items were built, from the beginning of the study, simultaneously in EP and BP by native speakers of each variety. Lexical adjustments were made whenever necessary for each Portuguese variety, but changes were kept as minimal. All the adjustments made after the pre-test, taking into consideration the participants' comments, were also made simultaneously for the two varieties. The sentences adjusted after the pre-test were not tested for their acceptability again. Therefore, the reported acceptability rates correspond to the sentences as they were presented in the pre-test prior to the final adjustments.

2.1.2 Procedure

Stimuli were presented in a self-paced reading task with a moving window (Just et al., 1982), where sentences were divided into segments of words or larger units, and the participants pressed a key to display each segment individually and proceeded at their natural reading pace. The time between the key presses was recorded and served as an indicator of participants' reading time. We applied a non-cumulative "moving window" version, replacing upcoming and previous segments with underscores (___), preserving white spaces.

Sentences were presented in PsychoPy–Pavlovia (Peirce et al., 2019), a web-based platform, segment-by-segment (segmentation is illustrated by the slashes (/) in example (36).

(36) *A Ana / garantiu / que / o Bruno / durante a visita / se cortou / a si mesmo / com o canivete / do jardim.*
"Anne / assured / that / Bruno / during the visit / SELF cut / himself / with the knife / of the garden."

The experimental items were pseudo-randomized with 64 filler items, with each experimental sentence always presented after at least one filler sentence. The experimental sentences were distributed in a Latin Square design. Hence, each participant read all the experimental sentences and was exposed to all the experimental conditions, but read only one version of each item, therefore, never reading the same sentence twice.

After each sentence (i.e., experimental and filler), a yes/no comprehension question was presented, where participants had to press the "N" key to answer "no" and the "S" key to answer "yes." For experimental sentences, the question always concerned

3 Available at https://onexp.textstrukturen.uni-goettingen.de/.

whether either the antecedent or the distractor was the subject of the complement clause [example (37a)] or not [example (37b)].

(37) a. *Foi o Bruno que se cortou?*
b. *Foi a Ana que se cortou?*
"Was it Bruno/Ana who cut himself?"

Half of the questions had "yes" as the correct answer [and thus asked about the antecedent in experimental sentences, *Bruno* in (37a)], and half had "no" as the correct answer [and asked about the distractor in experimental sentences, *Ana*, in (37b)]. The length of the questions-sentences was controlled [Median: 27 (25–28)].

The filler items were also complex sentences with different types of structures. To create some symmetry between the filler and the experimental items, we included sentences with two nouns with the same or different gender, as in the experimental items, but with no reflexive pronouns. We also included sentences with reflexive pronouns but in contexts with only one preceding referent. The length of the filler items was also similar to the experimental items. The filler items questions focused on different regions of the sentence but never focused on the verb of the subordinate clause. Again, half of the fillers had a "yes" answer, and the other half had a "no" answer, balancing "yes" and "no" between correct and incorrect responses. The fillers included sentences with two nouns, similar to the experimental items, that prompted "yes" answers as a correct response to questions triggering the first noun and "no" answers triggering the second noun (the opposite pattern of the experimental items). Filler questions allowed for a validation of reading for comprehension.

Before the experimental session, the participants went through a practice session with five training sentences to get used to the task.

2.2 Analysis and data preparation

Data preparation and analysis were conducted in R (version 4.2.3, R Core Team, 2023). We used the packages lme4 (version 1.1-34, Bates et al., 2015), and lmerTest (version 3.1-3, Kuznetsova et al., 2017), for the statistical analysis, and ggplot2 (version 3.5.0) for data visualization (Wickham, 2016).

Before the analysis, we excluded every trial in which reading times were lower than 200 ms and higher than 5,000 ms (see, for instance, Paape and Vasishth, 2022). Considering the length of each segment, a reading time higher than 5,000 ms is not realistically related only to the reading process and might reflect, instead, some distraction of the participant. Also, 200 ms is not enough time to process information during reading. Response times on questions lower than 1,000 ms or higher than 10,000 ms were also excluded.

2.2.1 Statistical models

For the statistical analysis, multilevel linear mixed models were built for the offline and online measures. As offline measures, we analyzed the response accuracy and response times for the question presented after the sentences, i.e., the time participants took to answer the final question. For the response time analysis, we only included the correct answers.

As an online measure, we analyzed reading times, that is, the time participants took to read each segment (from the beginning of the presentation of each segment until participants pressed a key to see the next segment).

We analyzed reading times in five regions [illustrated in (38)]: the pre-critical region, the critical region, the reinforcement region, the post-critical region, and the wrap-up region. Although both the gender-unmarked reflexive and the gender-marked reinforcement form are considered critical regions, for simplicity reasons, we use "critical" to refer to the region of the gender-unmarked and verb region. The pre-critical region precedes the gender-unmarked reflexive, and this region was analyzed to check if the impact of the Referents' Gender could be observed immediately after reading the two referents. The critical region includes the gender-unmarked reflexive and the verb (length: 9–10 characters). The reinforcement region corresponded to the segment where the gender-marked reflexive reinforcement form was presented (length: 10 characters), which occurs only in the ReflexReinf conditions (or ReinfOnly in BP). The post-critical region includes the segments posterior to the critical region in the ReflexOnly conditions and after the reinforcement region in the ReflexReinf conditions (length: 14–15 characters). The wrap-up region is the final region of the sentence (length: 8–12 characters). Linear mixed-effects models were computed separately for each region.

(38) A Ana garantiu que o Bruno [durante a viagem]$_{\text{Pre-Critical}}$ [se cortou]$_{\text{Critical}}$ [a si mesmo]$_{\text{Reinforcement}}$ [com o canivete]$_{\text{Post-Critical}}$ [do jardim.]$_{\text{Wrap-up}}$

All models included Referents' Gender as a main effect and as by-participants and by-items random slopes, as well as participants and items as random intercepts. Reflexive Type and its interaction with Referents' Gender were also included as main effects and as by-items and by-participants random slopes, when appropriate for the design, that is, in the post-critical and the wrap-up regions. Since the effect of Reflexive Type might only impact from the reinforcement region onward, before that region, Reflexive Type is not relevant for the analysis of on-line measures.

Following Laurinavichyute et al. (2017) (for further details, see Nicenboim et al., 2015), we included in the model participants' mean accuracy on the questions of the experimental items. This variable was scaled using the formula presented in Lago and Veríssimo (2023)'s script and was included only as a fixed effect.

We used sum-code contrasts and, in all models, for the Referents' Gender condition, the No-Overlap level was coded as −0.5, while the Overlap level was coded as +0.5. For the Reflexive Type, the ReflexReinf (or the ReinfOnly in BP) level was coded as −0.5, while the ReflexOnly level was coded as +0.5.

We used a linear mixed-effect model with a logistic link function for binomial answers (correct/incorrect) to analyze the response accuracy. Values were log-transformed to analyze the response times to the questions and the reading times. We always started with the full model and simplified it until convergence issues were solved. Nonetheless, all predictors are included as random slopes, although, sometimes, with no interaction between them.

2.3 Experiment 1: Reflexive pronoun resolution in European Portuguese

In this section, we describe the experiment we conducted for reflexive pronoun resolution in EP, contrasting two variables with two levels each: Referents' Gender (Overlap and No-Overlap) and Reflexive Type (gender-unmarked reflexive and gender-unmarked reflexive plus gender-marked reflexive, ReflexOnly and ReflexReinf conditions, respectively). We present a complete set of examples below.

(39) *O João garantiu que o Bruno durante a visita se cortou com o canivete do jardim.*

(40) *A Ana garantiu que o Bruno durante a visita se cortou com o canivete do jardim.*

(41) *O João garantiu que o Bruno durante a visita se cortou a si mesmo com o canivete do jardim.*

(42) *A Ana garantiu que o Bruno durante a visita se cortou a si mesmo com o canivete do jardim.*
"John / Anne assured that Bruno during the visit SELF cut (himself) with the knife of the garden."

2.3.1 Participants

Eighty-four participants were included in the analysis (female = 65, male = 18, and no answer = 1). The mean age of the participants was 25 years old (range: 19–56). Ten additional participants completed the online experiment, but their data were excluded due to speaking other languages (in addition to EP) at home (n = 7) or having an accuracy rate lower than 75% in filler sentences (n = 3). Participants were recruited by email and gave their informed consent before participating voluntarily in this experiment.

2.3.2 Analysis and data preparation

The cleaning criteria led to the exclusion of 8.7% of the data (5% for reading times on different segments and 3.7% for response times).

2.3.3 Results: offline measures

Overall, participants' mean accuracy was very high, at 91% (range: 69%–100%), as shown in Table 2. Nonetheless, the statistical analysis (Table 3) revealed a significant main effect of Referents' Gender, with lower accuracy in the Overlap condition, and of Type of Reflexive, with lower accuracy in the ReflexReinf condition.

For the response time of the correct answers, there is an effect of Referents' Gender (Table 4), with slower response times in the Overlap condition (Table 2).

2.3.4 Results: online measures

In the online measures, presented in Figure 1 and in Table 5, no effects were found for the pre-critical or the reinforcement regions. There are, however, effects of Referents' Gender in the critical and the post-critical regions, with slower reading times in the Overlap condition. There are also effects of Reflexive Type in the post-critical and the wrap-up regions, with slower reading times in the Overlap conditions.

There were no effects of Participants' Accuracy on any region, nor any interaction of Participants' Accuracy with any other factor.

TABLE 2 Experiment 1 (EP): Proportions of correct answers and mean response time.

Referents' gender	Reflexive type	Mean accuracy (SD)	Mean Resp. time (SD)
Overlap	ReflexOnly	0.878 (0.328)	2,533.053 (808.265)
Overlap	ReflexReinf	0.827 (0.378)	2,578.316 (997.602)
No-Overlap	ReflexOnly	0.967 (0.178)	1,941.742 (638.419)
No-Overlap	ReflexReinf	0.969 (0.174)	1,891.225 (563.278)

2.4 Discussion

In this experiment, we found effects of Referents' Gender on question-answering, with lower accuracy and slower response times in the Overlap conditions. In the offline measures, there was also an effect of Reflexive Type on the response accuracy, with lower accuracy rates on the ReflexReinf conditions, that is, on sentences with the gender-unmarked reflexive followed by the gender-marked reinforcement form.

In the online measures, effects of Referents' Gender were found in the critical and post-critical regions. Reading times were always slower in the Overlap condition.

Combining the online and offline results, the effects of the Referents' Gender in this experiment show evidence in favor of similarity-based encoding interference. When there is an overlap between the reflexive pronouns' antecedent and the distractor, whether the pronoun is gender-marked or gender-unmarked, there is a slowdown in reading speed, revealing higher processing costs and slower and less accurate responses.

On the other hand, there are no effects on the gender-marked reinforcement region. While the lack of effects might suggest an absence of interference effects triggered by the gender-marked reinforcement form, spill-over effects, similar to the ones found in the post-critical region, would be expected in the reinforcement region as well, since it is in between the two regions in which effects were detected. In the next paragraphs, we present possible explanations for this lack of results.

First, it is important to stress that this is a null result, meaning it needs to be interpreted with caution. This is even more important considering that the analysis in this region includes only half of the observations because this region exists only on half of the tested items (all sentences had gender-unmarked reflexive pronouns, but only half of them had the gender-marked reinforcement form).

However, an alternative explanation should be considered if the lack of effects might be interpreted as a null effect, that is, as an absence of evidence of interference. One possible answer is to consider that the absence of results is a consequence of the position of this element. As explained in the predictions of the

TABLE 3 Experiment 1 (EP): model results for accuracy on question-answering.

	Est	SE	*z*	*p*	CI (2.5%; 97.5%)
(Intercept)	3.040	0.154	19.737	<0.001	(2.738; 3.342)
RefsGender	−2.086	0.264	−7.898	<0.001	(−2.604; −1.568)
ReflexType	0.540	0.221	2.441	0.015	(0.106; 0.973)
RefsGender × ReflexType	0.323	0.406	0.795	0.427	(−0.473; 1.118)

TABLE 4 Experiment 1 (EP): Model results for response time for question-answering.

	Est	SE	df	*t*	*p*	CI (2.5%; 97.5%)
(Intercept)	7.594	0.026	87.979	289.730	<0.001	(7.543; 7.646)
RefsGender	0.234	0.025	44.676	9.323	<0.001	(0.185; 0.284)
ReflexType	0.010	0.019	44.032	0.538	0.593	(−0.027; 0.048)
RefsGender × ReflexType	−0.018	0.038	49.643	−0.467	0.642	(−0.093; 0.057)

Response times were log-transformed.

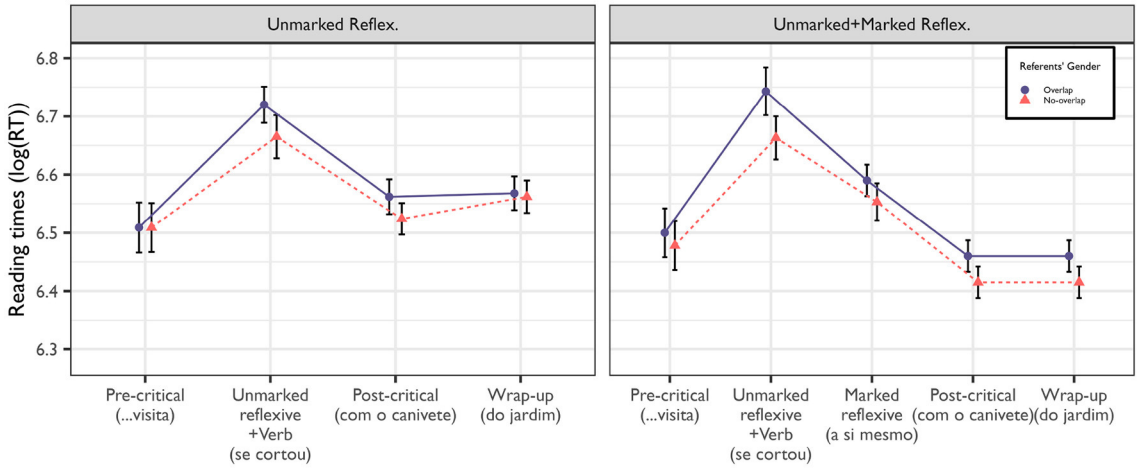

FIGURE 1
Experiment 1 (EP): Reading times in the different regions of analysis. Reading times were log-transformed.

study, the gender-marked reinforcement form, in this experiment, is always presented after the gender-unmarked reflexive pronoun. Therefore, when the gender-marked reinforcement form is read, its antecedent was already retrieved once and, thus, has increased its base-level activation, leading to a faster retrieval process. A similar lack of result is reported in Laurinavichyute et al. (2017) for the gender-marked reflexive in two different experiments, one in which the reflexive follows the verb, similar to the one reported here, and another in which it precedes the verb. The authors propose, after considering different possibilities, that the gender-marked reflexive, having an emphatic discursive function, might require additional semantic processing, which might hide the interference effects. These extra-processing costs might also hide the spill-over effect from the critical region, leading to no differences between overlapping and non-overlapping conditions. We get back to this discussion in Experiment 2.

When contrasting conditions with different types of reflexives, we found effects of Reflexive Type, with slower reading times in the conditions with only the gender-unmarked reflexive pronoun, both in the post-critical and in the wrap-up regions. This effect might be, however, an artifact of the experiment since there was a misalignment of the post-critical and the wrap-up regions in the different conditions of Reflexive Type: While in the ReflexOnly conditions, they corresponded to segments 7 and 8, respectively, in the ReflexReinf conditions, they corresponded to segments 8 and 9. This might be, therefore, an effect of a relation between reading speed and sentence length: reading speed might decrease as the sentence increases and as less relevant information is presented (in our experiments, the final segments were not crucial for the comprehension of the sentence). Also, as pointed out by one of the reviewers, this might also reflect spill-over effects of different syntactic categories. While in the ReflexOnly conditions, the post-critical region is preceded by a verb, a lexical category, in the ReflexReinf, it is preceded by a reflexive pronoun, which is likely to be read faster, causing less spill-over effects.

10.3389/flang.2024.1473948

TABLE 5 Experiment 1 (EP): Model results for the different regions of analysis.

	Est	SE	df	*t*	*p*	CI (2.5%; 97.5%)
Pre-critical region						
(Intercept)	6.498	0.040	83.284	163.471	<0.001	(6.420; 6.576)
RefsGender	0.015	0.021	30.189	0.695	0.493	(−0.027; 0.057)
AccPart	0.482	0.581	81.746	0.830	0.409	(−0.656; 1.621)
RefsGender × AccPart	−0.196	0.254	2287.325	−0.770	0.441	(−0.695; 0.303)
Critical region						
(Intercept)	6.699	0.038	101.400	174.284	<0.001	(6.624; 6.775)
RefsGender	0.066	0.019	30.320	3.441	0.002	(0.029; 0.104)
AccPart	0.257	0.498	81.270	0.517	0.607	(−0.718; 1.233)
RefsGender × AccPart	0.058	0.262	1194.000	0.221	0.825	(−0.456; 0.572)
Reinforcement region						
(Intercept)	6.572	0.027	77.700	239.077	<0.001	(6.518; 6.626)
RefsGender	0.031	0.024	231.400	1.282	0.201	(−0.017; 0.079)
AccPart	0.004	0.393	79.510	0.010	0.992	(−0.766; 0.774)
RefsGender × AccPart	0.281	0.357	470.300	0.788	0.431	(−0.418; 0.980)
Post-critical region						
(Intercept)	6.492	0.024	86.426	266.214	<0.001	(6.444; 6.540)
RefsGender	0.038	0.018	50.858	2.086	0.042	(0.002; 0.074)
ReflexType	0.110	0.023	39.709	4.872	<0.001	(0.066; 0.154)
AccPart	0.028	0.346	80.926	0.081	0.935	(−0.650; 0.706)
RefsGender × ReflexType	−0.006	0.033	53.667	−0.179	0.859	(−0.071; 0.059)
RefsGender × AccPart	0.403	0.260	79.744	1.551	0.125	(−0.106; 0.912)
ReflexType × AccPart	0.011	0.268	82.925	0.040	0.968	(−0.514; 0.535)
RefsGender × ReflexType × AccPart	0.652	0.473	166.695	1.378	0.170	(−0.275; 1.579)
Wrap-up region						
(Intercept)	6.526	0.029	99.622	222.620	<0.001	(6.469; 6.584)
RefsGender	0.025	0.017	51.911	1.537	0.130	(−0.007; 0.058)
ReflexType	0.079	0.019	36.952	4.204	<0.001	(0.042; 0.116)
AccPart	−0.259	0.382	80.887	−0.679	0.499	(−1.008; 0.489)
RefsGender × ReflexType	−0.046	0.030	218.121	−1.515	0.131	(−0.105; 0.013)
RefsGender × AccPart	−0.170	0.235	75.478	−0.726	0.470	(−0.630; 0.289)
ReflexType × AccPart	−0.352	0.235	109.850	−1.499	0.137	(−0.811; 0.108)
RefsGender × ReflexType × AccPart	0.410	0.449	223.560	0.913	0.362	(−0.470; 1.289)

Reading times were log-transformed.

Considering the acceptability judgment task, we would expect slower reading times in the condition with the lower acceptability rate, that is, the condition with both the gender-unmarked reflexive and the gender-marked reinforcement form. That was not the case, although we cannot easily explain the results, as mentioned in the previous paragraph. Still, it might reflect that participants' conscious judgments might differ from what they do during online and less controlled language processing. Actually, this is somewhat evident in the offline measures, in which we found lower accuracy rates in the Overlap conditions with the gender-marked reinforcement form. However, the design of our experiment does not allow us to present a clear explanation of the online results; therefore, further research is needed in this regard.

In sum, this experiment presents evidence of similarity-based encoding interference during gender-unmarked reflexive pronoun resolution and no evidence of retrieval interference. Moreover, the results highlight differences between the types of reflexive conditions tested in this experiment, with the online results but not the offline ones, contradicting our expectations. Considering the results of the acceptability judgment task, in which the reinforcement conditions had a low acceptability rating, slower reading times could be expected in these conditions. We return to this in the discussion of the next experiment.

2.5 Experiment 2: Reflexive pronoun resolution in Brazilian Portuguese

In this section, we describe the experiment we conducted for reflexive pronoun resolution in BP, contrasting two variables with two levels each: Referents' Gender (Overlap and No-Overlap) and Reflexive Type (gender-unmarked reflexive and gender-marked reflexive, ReflexOnly and ReinfOnly conditions, respectively). We present a complete set of examples above.

(43) *O João garantiu que o Bruno durante a visita se cortou com o canivete do jardim.*

(44) *A Ana garantiu que o Bruno durante a visita se cortou com o canivete do jardim.*

(45) *O João garantiu que o Bruno durante a visita cortou a si mesmo com o canivete do jardim.*

(46) *A Ana garantiu que o Bruno durante a visita cortou a si mesmo com o canivete do jardim.*
"John / Anne assured that Bruno during the visit (SELF) cut (himself) with the knife of the garden."

2.5.1 Participants

Ninety-six participants were included in the analysis (female = 49, male = 45, and no answer = 2). The mean age of the participants was 27 years old (range: 18–50). Thirteen additional participants completed the online experiment, but their data were excluded due to speaking another language (in addition to BP) at home (n = 6) or having an accuracy rate lower than 75% in filler sentences (n = 6). One participant was excluded as they were younger than 18 years. Participants were recruited by email or through social media and gave their informed consent to participate voluntarily in this experiment.

2.5.2 Analysis and data preparation

The cleaning criteria led to the exclusion of 10.9% of the data (5.8% for reading times on different segments and 5.1% for response times).

2.5.3 Results: offline measures

Overall, participants' mean accuracy was very high, with a mean of 90% (range: 66%–100%), as illustrated in Table 6. Nonetheless, the statistical analysis (Table 7) revealed a significant main effect of Referents' Gender, with lower accuracy in the Overlap condition.

For the response times of the correct answers, there is an effect of Referents' Gender (Table 8), with slower response times in the Overlap condition (Table 6).

TABLE 6 Experiment 2 (BP): Proportions of correct answers and mean response time by conditions.

Referents' gende	Reflexive type	Mean accuracy (SD)	Mean Resp. time (SD)
Overlap	ReflexOnly	0.846 (0.361)	2,453.063 (933.086)
Overlap	ReinfOnly	0.827 (0.379)	2,429.525 (820.286)
No-Overlap	ReflexOnly	0.959 (0.198)	1,950.316 (646.421)
No-Overlap	ReinfOnly	0.967 (0.178)	1,945.773 (535.961)

2.5.4 Results: online measures

In the online measures, presented in Figure 2 and in Table 9, there is a marginal effect of Referents' Gender in the post-critical region, with participants being slower in the Overlap condition. The effects of Participants' Accuracy were significant in every analyzed region, except for the reinforcement region in which the effect is marginal. Less accurate participants were consistently faster than more accurate ones. Moreover, there is a marginal interaction effect of Referents' Gender and Participants' Accuracy, in the reinforcement region, with less accurate participants, who are always faster readers, being slower in the No-Overlap condition. On the contrary, more accurate readers were slower in the Overlap condition.

There are also effects of Reflexive Type in the post-critical and the wrap-up regions, with participants being slower in the conditions with only the gender-unmarked reflexive. Moreover, in the wrap-up region, there is an interaction between Participants' Accuracy and Reflexive Type, which reflects a large difference between less accurate and faster readers and more accurate but slower readers in the conditions with only the reinforcement form (see Figure 3).

We conducted an additional analysis for the critical region. In this region, we had either the gender-unmarked reflexive and the verb or just the verb. Therefore, we were interested in this contrast as well, although this was not the focus of the study. The idea was to contrast the region with just the verb and with the verb and its complement, which was, at the same time, referentially identical to the subject of the verb. Therefore, differences were expected in this region. However, again, only effects of accuracy were found (see Table 10).

2.6 Discussion

In this experiment, we found effects of the Referents' Gender on accuracy and response times on question-answering with lower accuracy and slower response times in the Overlap condition. An effect of Participants' Accuracy was found in the analysis of every

TABLE 7 Experiment 2 (BP): Model results for Accuracy on question-answering.

	Est	SE	*z*	*p*	CI (2.5%; 97.5%)
(Intercept)	2.919	0.156	18.735	<0.001	(2.613; 3.224)
RefsGender	−2.269	0.214	−10.597	<0.001	(−2.689; −1.850)
ReflexType	0.031	0.196	0.159	0.874	(−0.354; 0.416)
RefsGender × ReflexType	0.287	0.361	0.794	0.427	(−0.421; 0.994)

TABLE 8 Experiment 2 (BP): Model results for Response Time for question-answering.

	Est	SE	df	*t*	*p*	CI (2.5%; 97.5%)
(Intercept)	7.591	0.026	101.795	293.172	<0.001	(7.540; 7.641)
RefsGender	0.189	0.018	40.558	10.573	<0.001	(0.154; 0.224)
ReflexType	−0.008	0.018	34.528	−0.417	0.679	(−0.044; 0.028)
RefsGender × ReflexType	0.017	0.034	52.734	0.496	0.622	(−0.049; 0.083)

Response times were log-transformed.

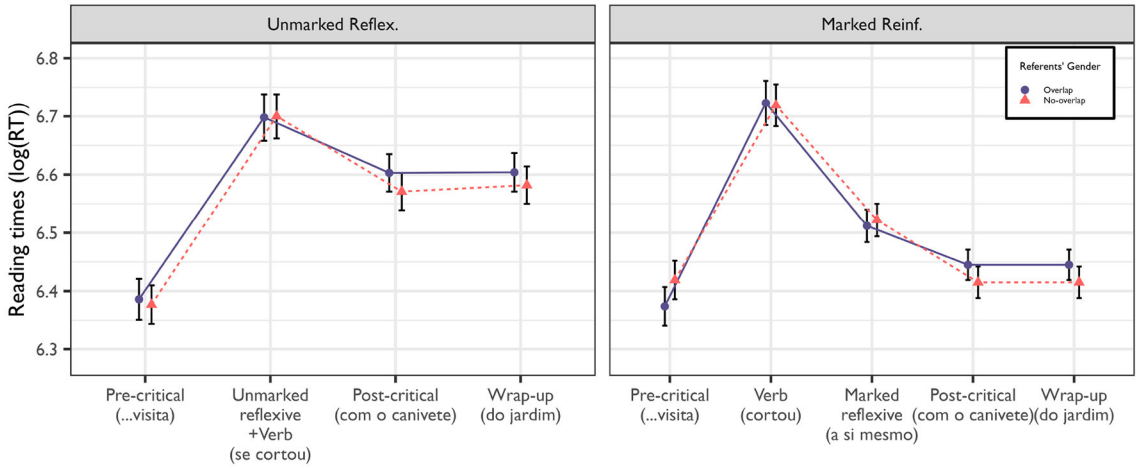

FIGURE 2
Experiment 2 (BP): Reading times in the different regions of analysis. Reading times were log-transformed.

region except for the reinforcement region. However, in the post-critical region, we found a marginal effect of Referents' Gender, with slower reading times in the Overlap condition. Moreover, in the reinforcement region, there was a marginal interaction between Participants' Accuracy and Referents' Gender, with less accurate readers being slower in the No-Overlap condition and the more accurate ones being slower in the Overlap one.

The interaction between Participants' Accuracy and Referents' Gender in the reinforcement region does not align with our predictions since it seems to show a facilitatory effect in the Overlap condition for the less accurate and faster readers. A similar pattern is also visible in the gender-unmarked reflexive region, when there was a gender-unmarked reflexive pronoun preceding the verb, although not statistically significant. Therefore, the lack of effects in the regions with the reflexives, both gender-unmarked and gender-marked, might be hidden by the opposite behavior of the less accurate readers and the more accurate ones. While the less accurate readers show a faster reading time in the Overlap condition, the more accurate readers show a slower reading time in the same condition. These opposite effects might have canceled each other out.

Additionally, while, at first sight, the facilitatory effect could be related to the low acceptability rate of the gender-marked reinforcement sentences found the pre-test, similar to what is described for ungrammatical sentences, the fact that the same pattern is also present in the gender-unmarked reflexive region contradicts that possibility. Therefore, a better explanation seems to be a shallow processing strategy adopted by the less accurate readers that might actually reflect an ambiguity advantage (Creemers and Meyer, 2022). Further analyses are, however, needed to clarify the reported results.

Since no effects, in addition to the interference reported in the previous paragraph, were found in the regions preceding the post-critical region, it is difficult to explain the source of the marginal effect of Referents' Gender found in the post-critical region. It

TABLE 9 Experiment 2 (BP): Model results for the different regions of analysis.

	Est	SE	df	*t*	*p*	CI (2.5%; 97.5%)
Pre-critical region						
(Intercept)	6.390	0.030	94.710	214.174	<0.001	(6.331; 6.448)
RefsGender	−0.020	0.015	526.265	−1.276	0.203	(−0.050; 0.010)
AccPart	1.123	0.389	93.967	2.886	0.005	(0.360; 1.885)
RefsGender × AccPart	0.039	0.202	2,074.462	0.194	0.846	(−0.357; 0.436)
Critical region						
(Intercept)	6.703	0.042	106.300	161.410	<0.001	(6.621; 6.784)
RefsGender	0.001	0.023	708.800	0.041	0.968	(−0.045; 0.047)
AccPart	1.116	0.474	93.710	2.355	0.021	(0.187; 2.044)
RefsGender × AccPart	0.249	0.311	984.900	0.799	0.425	(−0.362; 0.859)
Reinforcement region						
(Intercept)	6.524	0.025	84.171	257.432	<0.001	(6.474; 6.573)
RefsGender	−0.017	0.023	90.285	−0.725	0.471	(−0.062; 0.029)
AccPart	0.580	0.327	93.629	1.771	0.080	(−0.062; 1.221)
RefsGender × AccPart	0.562	0.306	93.956	1.837	0.069	(−0.038; 1.162)
Post-critical region						
(Intercept)	6.510	0.027	104.200	239.447	<0.001	(6.457; 6.563)
RefsGender	0.031	0.017	38.170	1.828	0.075	(−0.002; 0.063)
ReflexType	0.152	0.022	59.450	6.839	<0.001	(0.109; 0.196)
AccPart	0.787	0.341	93.960	2.306	0.023	(0.118; 1.456)
RefsGender × ReflexType	0.003	0.030	89.770	0.109	0.913	(−0.055; 0.061)
RefsGender × AccPart	−0.102	0.184	1,318.000	−0.554	0.580	(−0.462; 0.258)
ReflexType × AccPart	0.088	0.248	95.650	0.354	0.724	(−0.399; 0.575)
RefsGender × ReflexType × AccPart	0.410	0.369	1012.000	1.114	0.266	(−0.312; 1.133)
Wrap-up region						
(Intercept)	6.574	0.031	102.116	214.807	<0.001	(6.514; 6.634)
RefsGender	0.015	0.018	30.372	0.809	0.425	(−0.021; 0.051)
ReflexType	0.044	0.015	64.547	2.966	0.004	(0.015; 0.074)
AccPart	1.148	0.388	93.617	2.958	0.004	(0.387; 1.909)
RefsGender × ReflexType	0.026	0.029	713.460	0.922	0.357	(−0.030; 0.082)
RefsGender × AccPart	−0.184	0.197	104.100	−0.936	0.351	(−0.570; 0.202)
ReflexType × AccPart	−0.424	0.189	133.123	−2.241	0.027	(−0.795; −0.053)
RefsGender × ReflexType × AccPart	−0.094	0.379	735.397	−0.248	0.804	(−0.836; 0.648)

Reading times were log-transformed.

could be caused by increased processing costs either in the gender-unmarked region or in the gender-marked reinforcement region. Spill-over effects after the reinforcement region could be explained by encoding or retrieval interference, or both, caused by the gender-marked reinforcement reflexive form, while spill-over effects after the gender-unmarked reflexive would be explained by encoding interference only. Considering that the less accurate readers were faster in the Overlap condition in the preceding regions and that this effect might have hidden the interference effects of the more accurate readers, it is not possible to disentangle the source of the effect detected in the post-critical region.

Additionally, we also consider it relevant to stress the lack of effects in the critical region when contrasting the region with the reflexive plus the verb with the region with just the verb. Considering the amount of information provided on each region, we could expect differences between the two conditions. While in the Reflexive Only condition, more and more relevant information is provided (*Who did what to whom/what.*), in the Reinforcement

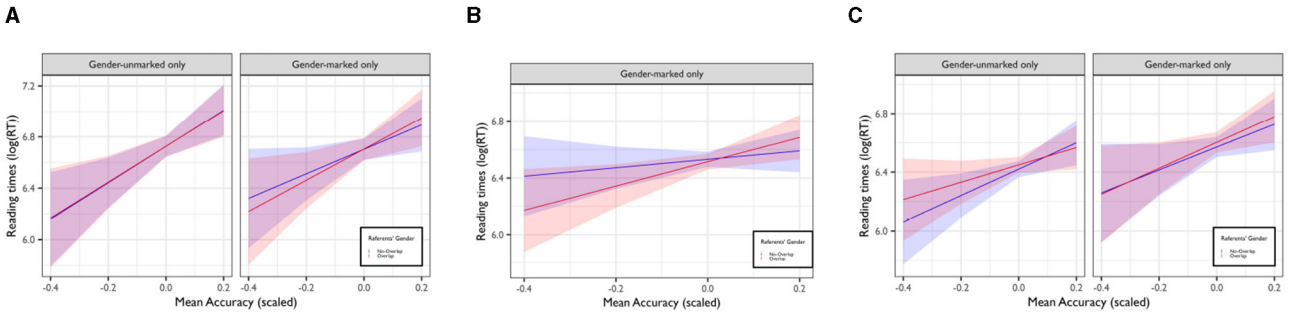

FIGURE 3
Experiment 2 (BP): Linear mixed model predicted values of reading times and participants' accuracy. Reading times were log-transformed, and mean accuracy was scaled. **(A)** Predicted values for the critical region. **(B)** Predicted values for the reinforcement region. **(C)** Predicted values for the post-critical region.

TABLE 10 Experiment 2 (BP): Results of the full model for the critical region.

Critical region (full model)						
	Est	SE	df	*t*	*p*	CI (2.5%; 97.5%)
(Intercept)	6.715	0.039	112.600	171.245	<0.001	(6.638; 6.792)
RefsGender	0.001	0.017	32.590	0.052	0.959	(−0.033; 0.035)
ReflexType	−0.022	0.020	29.330	−1.074	0.292	(−0.062; 0.018)
AccPart	1.248	0.443	93.800	2.818	0.006	(0.380; 2.116)
RefsGender × ReflexType	0.002	0.034	342.300	0.045	0.964	(−0.065; 0.068)
RefsGender × AccPart	0.121	0.226	750.900	0.538	0.591	(−0.321; 0.564)
ReflexType × AccPart	−0.325	0.233	93.970	−1.394	0.167	(−0.781; 0.132)
RefsGender × ReflexType × AccPart	0.271	0.445	1165.000	0.610	0.542	(−0.601; 1.144)

Reading times were log-transformed.

Only condition, only the verb (a transitive verb) is presented. Therefore, participants need to continue to the following region to get essential information to interpret the sentence. However, there are no effects, and a visual inspection allows us to see a lot of variability in the data (large SEs), leading to inconclusive results. The analysis of this region is of interest since it would allow us to contrast what happens when the subject is retrieved (also) as the direct object of the sentence with when the direct object is yet unknown. In the sentences we tested, in this region of the Reflexive Only conditions, the verb triggers the retrieval of the subject and of the object, which is the same entity. So, in that region, in the Reflexive Only conditions, two syntactic functions (and also two thematic roles) are attributed to the same entity, and, therefore, a processing load would be foreseen. Considering all this, encoding similarity-based interference would, thus, be expected and with different impacts on the two conditions. We consider that this is a topic that deserves further investigation in future work.

Finally, reading times were slower in the Reflexive Only condition, both in the post-critical and wrap-up regions. This effect was also found for EP despite the differences in the tested structures in the two varieties. Both in EP and in BP, reading times were slower in the final regions of the sentences with only the canonical gender-unmarked reflexive. These results seem to point toward facilitation in the conditions with the gender-marked reinforcement reflexive form. Nonetheless, as we previously mentioned, this might be an artifact of the experimental design, and, therefore, more data needs to be collected to understand what might be the explanation for this difference.

In this experiment, we consistently found effects of Participants' Accuracy on reading times, with less accurate readers being consistently faster than more accurate ones, showing that there is a trade-off between speed and accuracy. This result is in line with previous research (Nicenboim et al., 2015), and we consider that this effect, which reflects differences among participants, explains the overall results of the present experiment and, overall, the results of both experiments presented in the present study. While in the EP experiment there were no effects on Participants' Accuracy in any region, this effect was consistent in the BP experiment in all the regions of analysis. This result seems to indicate that there was more variability among the participants' performance in the BP experiment than in the EP one, which might explain why the effects are clearer in the EP experiment than in the BP one. We return to this in the next section.

3 General discussion

In the present study, we conducted two self-paced moving window reading experiments to test reflexive pronoun resolution in EP and BP. Our goal was to examine whether the reflexive's antecedent retrieval might be impacted by limitations imposed by memory antecedent's search, such as similarity-based interference, as proposed by cue-based retrieval models (Lewis and Vasishth, 2005), or if, instead, it is immune to this interference and is solely guided by syntactic constraints. We tested conditions with two referents of different or the same gender in sentences with gender-unmarked reflexive pronouns, gender-marked reinforcement reflexive forms, or both. The different types of structures allow us to test distinct reflexive properties in the same language, contrasting more directly the effects of encoding interference, retrieval interference, or both.

The offline results show lower accuracy and slower response times in the Overlap conditions, that is, when the two referents, the antecedent of the reflexive pronoun and the distractor, were of the same gender. These results were consistent across our two experiments, are constant across studies testing different languages (Jäger et al., 2015; Laurinavichyute et al., 2017), and are actually expected considering the model of Oberauer and Kliegl (2006): Encoding interference, caused by feature overwriting, might lead to more errors and higher processing times. Therefore, offline results point to encoding interference since accuracy was lower and response times were slower when there was gender overlap independently of the presence or absence of retrieval cues in the reflexive. However, offline results should be considered cautiously as evidence for encoding interference of reflexive pronoun resolution because they do not reflect the ongoing process of reflexive pronoun resolution and because offline and online results reflect different cognitive processes.

On the contrary, online results are not consistent across our experiments. Although we do find effects of gender overlap in the EP experiment, we only find marginal effects on the BP experiment in the post-critical region and an interaction effect of Referents' Gender and Participants' Accuracy in the reinforcement region.

Nevertheless, the inconsistency of our results is in line with the variability found in the different studies previously mentioned. If we consider the results from different studies, even higher-power ones (e.g., Jäger et al., 2015; Laurinavichyute et al., 2017), the results are hardly consistent. For instance, while Jäger et al. (2015) and Laurinavichyute et al. (2017) did not find effects of interference with German gender-unmarked reflexives, Laurinavichyute et al. (2017) found effects of encoding interference for the Russian gender-unmarked reflexive, but not for the gender-marked ones. Our results are also in line with previous studies, as the ones mentioned above, in what concerns the dissimilarity between inconsistent reading times results and robust offline accuracy ones.

Moreover, the inconsistency of results is actually only evident in studies focusing on grammatical structures. The studies that include ungrammatical conditions, in which the reflexive is forced to retrieve an ungrammatical antecedent (e.g., "Mary saw that Peter hurt herself."), consistently find facilitatory interference effects. This difference between grammatical and ungrammatical constructions and the lack of clear evidence of inhibitory interference effects in grammatical conditions was also stressed by Jäger et al. (2020) and Yadav et al. (2022), for instance. Moreover, and also as pointed out by Jäger et al. (2020), inhibitory interference effects might be subtle and complex to find. This seems to be clear from our data. If we take a closer look at the results of the models, it is clear that the effect of the Referents' Gender is always very subtle (in online reading times), while, at the same time, there is always large inter-individual variability.

Actually, inspecting the random structure of the models (see Table 11), it is very clear that reading speed varies a lot between the participants, and this variability seems to be stronger than the interference effect manipulated in our experiments. While the random intercept for participants is large, the random slope of Referents' Gender is small, showing little variability between the participants. However, the effect is also always very small when we look at the estimate of the model. For instance, in the critical region in the EP experiment, the Referents' Gender standard deviation by the participant is 0.020, while the effect estimate in the model is 0.066, and, in the post-critical, while the standard deviation by the participant is 0.085, the effect estimate of the model is 0.038. In the post-critical region of the BP experiment, in which the effect is marginally significant, the standard deviation by the participant is 0.015, and the model estimate is 0.031. So, it seems that the combination of large inter-individual variability on overall mean reading times, with a small effect of the Referents' Gender both in the models' estimates and as by-participant slopes, leads to a fading of the potential effect of similarity-based interference.

This perspective is corroborated by a model comparison. Following Lago and Veríssimo (2023) and Frinsel and Christiansen (2024), we conducted model comparisons for every model of our analysis. The results show consistent improvements when we include by-participant intercepts. So, it is clear that there is consistent individual variability in the overall reading speed in both experiments in every region of analysis. Moreover, comparing the models with and without Referents' Gender as random slope by-participants does not show any improvement, which reflects little variability between participants on the effect size of Referents' Gender.

Moreover, taking into account the fact that, in the BP experiment, contrary to what happened in the EP experiment, there were consistent effects of Participants' Accuracy on reading speed, and only marginal effects of Referents' Gender were found, it seems that the combination of large variability among participants with a small effect size might easily dilute the impact of this effect. Hence, larger sample-size studies are required, as reinforced by Jäger et al. (2020). Moreover, considering the interaction effect between Participants' Accuracy and Referents' Gender found in the BP experiment, we believe that future studies need to account for inter-individual variability, including also the perspective of deep and shallow processing strategies (e.g., Laurinavichyute, 2021; Creemers and Meyer, 2022).

Finally, we made predictions about the reflexives' antecedent retrieval at the different points of the sentence. We expected that, based on the cue-based retrieval model of Lewis and Vasishth (2005), similarity-based interference effects would be lower or even null in the reinforcement form in EP, since the antecedent was already retrieved once (immediately before), and is, therefore,

TABLE 11 Random structure of Critical and Post-critical regions of analysis from the two experiments: EP and BP.

		Critical		Post-critical	
		EP (SD)	BP (SD)	EP (SD)	BP (SD)
Part	(Intercept)	0.299	0.334	0.204	0.245
	RefsGender	0.020	0.024	0.085	0.015
	ReflexType			0.095	0.125
	RefsGender × ReflexType			0.107	0.034
Item	(Intercept)	0.103	0.117	0.037	0.045
	RefsGender	0.042	0.010	0.031	0.053
	ReflexType			0.075	0.067
	RefsGender × ReflexType			0.055	0.059
Residual		0.436	0.421	0.365	0.357

easier to retrieve (due to its higher level activation), or might even not be retrieved again. On the other hand, in BP, both the gender-unmarked and the gender-marked reflexives would trigger retrieval of the reflexive's antecedent for the first time. The results for EP align with this prediction since no effects were found in the gender-marked reinforcement region, while interference effects were found in the gender-unmarked reflexive. The BP results, in which we found a conflicting and unexpected result of the less accurate readers, nonetheless, might also be interpreted as showing retrieval activation at both the gender-unmarked reflexive and the gender-marked reinforcement form. Although the effects are not statistically significant, there seems to exist a difference between the Overlap and No-Overlap conditions at the reflexives site, both in the critical region, the gender-unmarked reflexive (compared to the conditions without reflexive form preceding the verb), and the reinforcement region, the gender-marked form. However, new experiments, specifically built to analyze this question and with the proper characterization of the participant's profile, need to be conducted to further explore this topic.

4 Conclusions

In the present study, we found inconsistent results across two experiments testing reflexive pronoun resolution in Portuguese. Although the offline results point toward similarity-based encoding interference effects during reflexive pronoun resolution in Portuguese, the online results only partially support this perspective. In addition, we suggest that inter-individual differences have a high impact on the results, as the interference effect is usually small and seems to fade whenever there is a large difference between participants' mean reading times or mean accuracy. Additionally, less accurate and more accurate participants might even show opposite reading strategies.

Our study contributes to the discussion of reflexive pronoun resolution, enlarging the set of tested languages. Results show encoding similarity-based interference during reflexive pronouns antecedents' retrieval. Additionally, with respect to the variability of results, we suggest that individual differences play a crucial role in the interference effect observed during online reflexive pronoun resolution of grammatical structures, although the topic requires further investigation. The analysis of individual differences, which is a growing topic of interest in psycholinguistics (Yadav et al., 2022), although not yet easily exploitable, might be a helpful tool in exploring the impact of similarity-based interference on reflexive pronoun resolution in the future.

Data availability statement

The datasets and the analysis scripts are available at the OSF repository: https://osf.io/jn2kz/.

Ethics statement

The studies involving humans were approved by School of Arts and Humanities Ethical Committee, University of Lisbon (7_CEI2018). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

PL: Conceptualization, Formal analysis, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing, Investigation, Visualization. ML: Conceptualization, Methodology, Writing – review & editing, Funding acquisition. DA-V: Writing – review & editing, Validation, Visualization. JG: Methodology, Writing – review & editing, Investigation, Visualization. AC: Conceptualization, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was financed by Fundação para a Ciência e a Tecnologia, with a Postdoctoral fellowship awarded to Paula Luegi (SFRH/BPD/84138/2012), by CAPES, with a Postdoctoral

fellowship awarded to Márcio Leitão (99999.005246/2014-08), and also by CNPq, with the *Produtividade em Pesquisa*—PQ 2015 (308578/2015-5) fellowship, awarded to Márcio Leitão. Daniela Avila-Varela was supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia with a Postdoctoral fellowship under the project UIDB/00214/2020 (https://doi.org/10.54499/UIDB/00214/2020) awarded to CLUL. Jéssica Gomes was supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia with a PhD fellowship (UI/BD/154500/2022).

Acknowledgments

We are thankful to the reviewers of the paper for their insightful comments and to all the volunteers who agreed to participate in our experiments. We also want to express our gratitude to Mário Carvalho and Matheus de Almeida, who worked on this project in its early steps. Finally, we would like to thank João Veríssimo for his insights about this study and for his valuable comments on a previous version of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Badecker, W., and Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 748–769. doi: 10.1037//0278-7393.28.4.748

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht, Foris.

Clackson, K., and Heyer, V. (2014). Reflexive anaphor resolution in spoken language comprehension: structural constraints and beyond. *Front. Psychol.* 5:904. doi: 10.3389/fpsyg.2014.00904

Creemers, A., and Meyer, A. S. (2022). The processing of ambiguous pronominal reference is sensitive to depth of processing. *Glossa Psycholinguist.* 1, 1–30. doi: 10.5070/G601166

Cunnings, I., and Felser, C. (2013). The role of working memory in the processing of reflexives. *Lang. Cogn. Process.* 28, 188–219. doi: 10.1080/01690965.2010.548391

Dillon, B. (2014). Syntactic memory in the comprehension of reflexive dependencies: an overview. *Lang. Linguist. Compass* 8, 171–187. doi: 10.1111/lnc3.12075

Dillon, B., Mishler, A., Sloggett, S., and Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *J. Mem. Lang.* 69, 85–103. doi: 10.1016/j.jml.2013.04.003

Fonseca, P. (2012). Os verbos pseudo-reflexivos em Português Europeu. *ElingUP: Rev. Eletrón. Linguíst. Estud. Univ. Porto* 4, 31–49.

Frinsel, F., and Christiansen, M. (2024). Capturing individual differences in sentence processing: how reliable is the self-paced reading task? *Behav. Res. Method* 56, 6248–6257. doi: 10.3758/s13428-024-02355-x

Gordon, P., Hendrick, R., and Levine, W. (2002). Memory-load interference in syntactic processing. *Psychol. Sci.* 13, 425–430. doi: 10.1111/1467-9280.00475

Jäger, L., Benz, L., Roeser, J., Dillon, B., and Vasishth, S. (2015). Teasing apart retrieval and encoding interference in the processing of anaphors. *Front. Psychol.* 6:506. doi: 10.3389/fpsyg.2015.00506

Jäger, L., Engelmann, F., and Vasishth, S. (2017). Similarity-based interference in sentence comprehension: literature review and Bayesian meta-analysis. *J. Mem. Lang.* 94, 316–339. doi: 10.1016/j.jml.2017.01.004

Jäger, L., Mertzen, D., Van Dyke, J., and Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: a large-sample study. *J. Mem. Lang.* 111:104063. doi: 10.1016/j.jml.2019.104063

Just, M. A., Carpenter, P. A., and Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *J. Exp. Psychol. Gen.* 111, 228–238. doi: 10.1037/0096-3445.111.2.228

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lago, S., and Veríssimo, J. (2023). *Tutorial on Analyzing Individual Variability in Language Processing*. Charlottesville, VA: OSF.

Laurinavichyute, A., Jäger, L., Akinina, Y., Roß, J., and Dragoy, O. (2017). Retrieval and encoding interference: cross-linguistic evidence from anaphor processing. *Front. Psychol.* 8:965. doi: 10.3389/fpsyg.2017.00965

Laurinavichyute, A. K. (2021). *Similarity-based interference and faulty encoding accounts of sentence processing*. (Dissertation). Potsdam: University of Potsdam.

Leitão, M., Brito, D. B., and Leite, N. (2017). "Formas anafóricas reflexivas em Português Brasileiro e Português Europeu: um estudo de corpora," in *XXV ENIC - Encontro de Iniciaç˜ao* Científica da UFPB (João Pessoa).

Lewis, R. L., and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 375–419. doi: 10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., and Dyke, J. A. V. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* 10, 447–454. doi: 10.1016/j.tics.2006.08.007

McElree, B., Foraker, S., and Dyer, L. (2003). Memory structures that subserve sentence comprehension. *J. Mem. Lang.* 48, 67–91. doi: 10.1016/S0749-596X(02)00515-6

Mello, F. R. (2009). *Acabou-se o que era doce, quem comeu se regalou-se: uma análise do clítico se em João Pessoa na interface Sociolinguística* (PhD dissertation). João Pessoa: Universidade Federal da Paraíba.

Menuzzi, S., and Lobo, M. (2016). "Binding and pronominal forms in Portuguese," in *The Handbook of Portuguese Linguistics*, eds. W. L. Wetzels, S. Menuzzi, and J. Costa (Oxford: Wiley-Blackwell), 338–355. doi: 10.1002/9781118791844.ch18

Nicenboim, B., Engelmann, F., Suckow, K., and Vasishth, S. (2015). "Fail fast or succeed slowly: Good-enough processing can mask interference effects," in *Proceedings of the 13th International Conference on Cognitive Modeling* (Groningen), 196–197.

Nicol, J., and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *J. Psycholinguist. Res.* 18, 5–19. doi: 10.1007/BF01069043

Oberauer, K., and Kliegl, R. (2006). A formal model of capacity limits in working memory. *J. Mem. Lang.* 55, 601–626. doi: 10.1016/j.jml.2006.08.009

Paape, D., and Vasishth, S. (2022). Is reanalysis selective when regressions are consciously controlled? *Glossa Psycholinguist.* 1, 1–34. doi: 10.5070/G601139

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, M. R., Kastman, H. S. E., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203. doi: 10.3758/s13428-018-01193-y

Pereira, D. C. (2007). *Variaç˜ao* e mudança no uso dos pronomes reflexivos no Português popular da capital Paulista (PhD dissertation). São Paulo: Universidade de São Paulo.

Phillips, C., Wagers, M., and Lau, E. (2011). "Grammatical illusions and selective fallibility in real-time language comprehension," in *Experiments at the*

Interfaces, Vol. 37, ed. J. Runner (Bingley: Emerald Publications), 147–180. doi: 10.1163/9781780523750_006

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* 48, 542–562. doi: 10.1016/S0749-596X(02)00536-3

Villata, S., Tabor, W., and Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: evidence from agreement. *Front. Psychol.* 9:2. doi: 10.3389/fpsyg.2018.00002

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. doi: 10.1007/978-3-319-24277-4_9

Yadav, H., Paape, D., Smith, G., Dillon, B. W., and Vasishth, S. (2022). Individual differences in cue weighting in sentence comprehension: an evaluation using approximate Bayesian computation. *Open Mind* 6, 1–24. doi: 10.1162/opmi_a_00052

10.3389/flang.2024.1473948

Appendix

Table A1 Experiment 1 (EP): Mean and standard deviation of the raw reading time for each region.

NPGender	ReflexCond	Segment	*M* (SD)
Overlap	ReflexOnly	Unmarked+Verb	802.514 (406.297)
Overlap	ReflexOnly	Post-critical	957.858 (278.739)
Overlap	ReflexOnly	Wrap-up	913.516 (346.578)
Overlap	ReflexReinf	Unmarked+Verb	755.606 (217.669)
Overlap	ReflexReinf	Marked	789.296 (229.166)
Overlap	ReflexReinf	Post-critical	788.358 (228.545)
Overlap	ReflexReinf	Wrap-up	811.949 (413.251)
No-Overlap	ReflexOnly	Unmarked+Verb	1,020.709 (453.654)
No-Overlap	ReflexOnly	Post-critical	914.685 (345.595)
No-Overlap	ReflexOnly	Wrap-up	829.654 (277.396)
No-Overlap	ReflexReinf	Unmarked+Verb	711.609 (218.104)
No-Overlap	ReflexReinf	Marked	669.048 (223.834)
No-Overlap	ReflexReinf	Post-critical	762.332 (273.93)
No-Overlap	ReflexReinf	Wrap-up	709.437 (221.442)

Table A2 Experiment 2 (BP): Mean and standard deviation of the raw reading time for each region.

NPGender	ReflexCond	Segment	*M* (SD)
Overlap	ReflexOnly	Unmarked+Verb	674.129 (262.761)
Overlap	ReflexOnly	Post-critical	961.931 (416.446)
Overlap	ReflexOnly	Wrap-up	954.982 (386.744)
Overlap	ReinfOnly	Unmarked	806.004 (292.919)
Overlap	ReinfOnly	Marked	828.488 (274.631)
Overlap	ReinfOnly	Post-critical	811.706 (279.23)
Overlap	ReinfOnly	Wrap-up	677.068 (266.22)
No-Overlap	ReflexOnly	Unmarked+Verb	992.274 (394.191)
No-Overlap	ReflexOnly	Post-critical	970.443 (359.319)
No-Overlap	ReflexOnly	Wrap-up	771.959 (267.538)
No-Overlap	ReinfOnly	Unmarked	695.21 (216.931)
No-Overlap	ReinfOnly	Marked	661.328 (201.498)
No-Overlap	ReinfOnly	Post-critical	796.838 (293.086)
No-Overlap	ReinfOnly	Wrap-up	795.773 (303.954)