




## ARTICLE

# Third-Person Object Anaphora in Brazilian Portuguese: Full Pronouns and Null Objects in the Speech of Salvador

## O Objeto Anafórico de Terceira Pessoa no Português Brasileiro: Pronomes Plenos e Objetos Nulos na Fala de Salvador

Cristina Figueiredo   
Lílian Teixeira de Sousa   
João Paulo Lazzarini-Cyrino 

Universidade Federal da Bahia, Salvador, BA, Brasil

E-mails: [macrisfig@uol.com.br](mailto:macrisfig@uol.com.br); [lilian.sousa@ufba.br](mailto:lilian.sousa@ufba.br); [jpcyrino@gmail.com](mailto:jpcyrino@gmail.com)

**ABSTRACT:** In Brazilian Portuguese (BP), the third-person object anaphora has important features: The clitic anaphora is disappearing, while the full pronoun and null object are becoming the norm. The distribution of the two strategies has been studied within sociolinguistics (Omena, 1978; Duarte, 1986; Figueiredo, 2004), generative theory (Raposo, 1986; Galves, 1989; Cyrino 1993, 1997), and diachronic linguistics (Cyrino, 1997; Macedo-Costa, 2012). These studies agree that animacy and specificity are related to the occurrence of the full pronoun strategy. Schwenter (2006, 2014), Cyrino (2017) propose that the variation may be due to Differential Object Marking. For Creus and Menuzzi (2004), the lack of semantic gender on the antecedent would favor the null object form. This study investigates these three aspects in the speech of individuals from three neighborhoods of Salvador/Bahia. We considered the extralinguistic factors: Age, schooling, and the neighborhood itself. Chi-squared tests revealed that 5 out of 6 variables are significant ( $p < .001$ ), with schooling being the one ruled out. Logistic Regression analysis showed the odds are mostly in favor of null objects as the chosen anaphoric strategy, with definite and indefinite non-specific antecedents and the Plataforma neighborhood as the main factors contributing to a higher odds ratio in favor of full.

**KEYWORDS:** Null Object, Full Pronoun, Animacy.

**RESUMO:** A realização do objeto direto anafórico no português brasileiro (PB) é um fenômeno variável, que pode apresentar um clítico acusativo, que está desaparecendo, o pronome pleno (PL) ou o objeto nulo (ON). Esse fenômeno tem sido analisado sob os pressupostos de perspectivas teóricas distintas: pela Sociolinguística (Omena, 1978; Duarte, 1986; Figueiredo, 2004), pela Teoria Gerativa (Raposo, 1986; Galves, 1989; Cyrino 1993, 1997) e pela Linguística Diacrônica (Cyrino, 1997; Macedo-Costa, 2012). Esses estudos apontam que a animacidade e a especificidade estão relacionadas à escolha das duas estratégias mais produzidas, ON e PL. Schwenter (2006, 2014),

### COMO CITAR

FIGUEIREDO, Cristina;  
TEIXEIRA DE SOUSA, Lílian;  
LAZZARINI-CYRINO, João  
Paulo. Third-Person Object  
Anaphora in Brazilian  
Portuguese: Full Pronouns and  
Null Objects in the Speech of  
Salvador. *Revista da Anpoll*, v.  
55, e1924, 2024. doi: <https://doi.org/10.18309/ranpoll.v55.1924>



Cyrino (2017) propõem que a variação é devida à Marcação Diferencial de Objetos, justificando a emergência do PL. Creus e Menuzzi (2004) acrescentam que a falta de gênero semântico do antecedente favoreceria o objeto nulo. Considerando esses fatores, neste artigo, apresenta-se o resultado da investigação da variação entre o objeto nulo e o pronome pleno, na fala de indivíduos de três bairros de Salvador/Bahia. Como fatores extralinguísticos, fixamos: faixa etária, escolaridade e bairro. A análise de regressão logística mostrou objetos nulos como estratégia anafórica escolhida. A probabilidade de escolha do pronome pleno é verificada quando o antecedente é definido ou indefinido não-específico e quando ocorre entre moradores do bairro de Plataforma.

**PALAVRAS-CHAVE:** Objeto nulo, Pronome pleno, Animacidade.

## 1 Introduction

Studies on Brazilian Portuguese (henceforth BP) have shown important syntactic phenomena that distinguish it from other varieties, being third-person object anaphora one of them. In BP, the clitic anaphora (*o/a/os/as*) is disappearing even in formal speech, while the full pronoun (*ele/ela/eles/elas*) and null object varieties are becoming the norm. The distribution of the two strategies has been studied within sociolinguistics, generative theory, and diachronic linguistics. These studies agree that animacy and specificity are related to the occurrence of the full pronoun strategy, but not categorically. Schwenter (2006, 2014) and Cyrino (2017), for instance, propose that the variation may be due to Differential Object Marking. Therefore, direct objects in BP could be marked differently due to the [+animate] feature present in full pronouns. Null objects also show up with animate antecedents, albeit in a lower frequency. For Creus and Menuzzi (2004), the relevance of animacy is related to the fact that the referents show a morphological distinction of semantic gender. Thus, the lack of semantic gender on the antecedent would favor the null object form. A recent study (Teixeira de Sousa and Figueiredo, 2022) analyzed the distribution of null objects and full pronouns regarding the semantic of the antecedents. Their *corpus*, however, was recorded in the 1980s, in an isolated community in the rural area of Bahia, majorly composed of Afro-Brazilians. However, there are no analyzes dealing with recent data from urban varieties from the same region to compare. Having that said, our aim was to carry out an exploratory study on this variation through data from the speech of individuals from three neighborhoods of Salvador/Bahia (Itapuã, Liberdade, and Plataforma), bearing in mind their socio-historical characteristics. Then, we investigated the relevance of semantic features (animacy, specificity, and semantic gender) to the decision between one strategy or the other. We also considered the extralinguistic factors: age group, educational level, and the neighborhood itself.

We conducted statistic tests as chi-squared and logistic regression. The Chi-squared tests revealed that 5 out of 6 variables are significant ( $p < .001$ ), with educational level being the one ruled out. Logistic Regression analysis showed the odds are mostly in favor of null objects, with definite and indefinite non-specific antecedents and the Plataforma neighborhood as the main factors contributing to a higher odds ratio in favor of full pronouns. These data are better discussed in our last section.

## 2 Direct Object Anaphora (DOA)

There are four different strategies for referring to a third person antecedent in Portuguese. They are unevenly distributed in our sample.

The third person **accusative clitic (CL)**, “o(s)” and “a(s)”, is generally infrequent in spoken BP, being mostly used in formal situations or in written language. They occur in post-verbal positions of infinitive clauses, having their syllabic onset filled by the “l” buffer consonant such as in *beijá-lo* (*kiss him*). They can also occur in gerund clauses and in a few simple SVO structures with simple tenses (Omena, 1978; Duarte, 1986; Arruda, 2006; among others). Our data showcased only three instances of it.

The **repetition of the antecedent DP and its further recovery by an anaphoric DP (DP)** receives neutral evaluation from the speakers. It is a replacement for the CL strategy and, as reported by Cyrino (1997), is increasingly becoming more infrequent. In our data, the DP strategy corresponds to 29% of the observed strategies.

The **null object (NO)**, is evaluated by the speakers as a neutral strategy and is very common in the colloquial (Figueiredo, 2004; 2006) and standard (Duarte, 1986; Arruda, 2006) varieties of BP. It is also commonly found in written Portuguese (Averbug, 2000; Othero; Schwanke, 2018; Othero et al., 2018). Linguists have frequently associated the NO strategy with antecedents bearing the [-animate] feature, as in (3a). However, we could find NO occurrences associated with both [+animate] (3b) and [-animate] antecedents in our sample. NOs correspond to 86% , of the strategies found in our data.

- (3) a. *dô o dinheiro.* *Cá-pôco ele vem: “Toma \_\_\_\_i af”.* (LIB04)  
 give.1sg the money soon he comes take Ø there  
 ‘I give him money. Soon he comes back and says: ‘Take it.’
- b. *o filho dela mora aqui com minha mãe.*  
 the son her lives here with my mother  
*Minha mãe trouxe \_\_\_\_i pra morá aqui...* (PLAT01)  
 my mother brought Ø for live here  
 ‘Her son lives here with my mother. My mother brought him to live here.’

Finally, **full pronouns (FP)** are an anaphoric strategy for referring to direct objects. These are traditionally considered subject pronouns and should refer to subjects rather than objects (*João<sub>i</sub> é meu amigo. Ele<sub>i</sub> está sempre a meu lado; João<sub>i</sub> is my friend. He<sub>i</sub> is always by my side.*) However, besides being able to refer to direct objects, they can also get case from a preposition (*João<sub>i</sub> é meu amigo. Dei a ele<sub>i</sub> um presente; John<sub>i</sub> is my friend. I gave him<sub>i</sub> a gift.*) Their usage in television programs, movie subtitles and even within journalistic texts on the internet suggests they might also occur among literate speakers. Differently from NOs, FPs are commonly associated with [+animate] antecedents. This is not categorical, however. In our data, FPs correspond to 10% of the observed strategies.

- a. *Ela<sub>i</sub> tem treze ano. Levo —<sub>i</sub> pa escola e vô buscá ela<sub>i</sub>*  
 she has thirteen year take-along Ø to school and go take she  
 ‘She is thirteen years old. I drive her to school and back home.’
- b. *fazia aquele saquinho, pegava areia da praia, enchia ele<sub>i</sub>*  
 do.1.sg.Pst.Prog that bag.M take.Pst.Prog sand of-the beach fill.Pst.Prog he  
 ‘I used to make a little bag, and fill it with sand from the beach.’

Considering the innovative characteristics of NOs and FPs, this paper presents a description of the behavior of both strategies in the popular (working class) urban Portuguese variety spoken in the city of Salvador, Bahia. We thereby attempt to associate animacy, specificity, and gender features with their diffusion in BP.

### 3 Null Objects and Full Pronouns: An Overview of Their History in BP

Many works published since the 1980s (Duarte, 1986; Cyrino, 1997) highlight the following aspects of BP ODA: i) The possibility of a nominative FP to occur in accusative positions, as shown in (4); ii) Usage of CL, as in (1), reserved to written language and, in spoken language, only to more formal contexts in the enclitic forms *lo(s)* and *la(s)*; and iii) Broader usage of NOs, even in syntactic contexts not allowed in European Portuguese, such as syntactic islands (Raposo, 1986).

Huang’s (1984) study on Chinese null objects has inspired a series of studies attempting to investigate, within the scope of the generativist framework, the status of the direct object gap in Brazilian Portuguese (variable, *pro* or ellipsis<sup>1</sup>): Galves (1984, 1989), Kato (2001), Cyrino (1993, 1997), Figueiredo (2009; 2012), Cyrino, Duarte, Kato (2020). Within diachronic linguistics, Cyrino (1997), Macedo-Costa (2012), and Figueiredo (2019) analyzed the occurrence of NOs and FPs seeking traces of the period in which these strategies emerged in BP. Under the sociolinguistics perspective, although DOA has already been studied by Omena (1978), it is Duarte (1986) who is more well-versed in this respect. Her study is particularly remarkable for determining important explanatory variables, such as animacy, for the occurrence of the different strategies.

According to Cyrino (1997), data from the 17<sup>th</sup> century suggested the empty category in the direct object position becomes robust evidence for BP acquirers. By the 19<sup>th</sup> century, through clausal ellipsis, with [+specific, -animate] features, a child acquiring the language would extend the possibility of DP ellipsis to anaphoric objects with the [+specific, -animate] features. Also, in the same century, FPs begin to be observed as direct object anaphora in direct object positions when their antecedent is [+animate]. In addition to those observations, Figueiredo (2019) analyses 19<sup>th</sup> century records from the Sociedade Protetora dos Desvalidos (Society for the Protection of the Helpless). The texts contained direct object FPs referring to

<sup>1</sup> For a thorough discussion on the status of the empty category in the direct object position, see Cyrino (1997) and Figueiredo (2009, 2012), among others.

[-animate] antecedents. Given the slower nature of written language with respect to linguistic change, it is believed that the occurrence of FPs as direct object anaphora referring to [-animate] antecedents might have entered BP's grammar anytime earlier than the 19<sup>th</sup> century.

Another attempt to explain the emergence of NOs and FPs as direct object anaphora comes from Figueiredo (2004). She hypothesized that the changes seen in BP were a result of linguistic contact, primarily between Portuguese and African languages during the colonial period. She goes on to analyze data from four Afro-Brazilian communities living in relative isolation. Surprisingly, NOs are most prevalent in the group composed of individuals over 60 years old, with 91% of the occurrences. The remaining 9% are FP occurrences. These findings are in line with previous works' findings — that [-animate] antecedents favor the occurrence of NO anaphora, whereas [+animate] antecedents favor FP anaphora. The specificity feature of the antecedent was also proved to be significant: [-specific] antecedents favor NO anaphora, whereas [+specific] antecedents favor FP anaphora. However, no context is categorical, and variation is observed among the communities. It is worth mentioning that the data from Afro-Brazilian communities show no occurrence whatsoever of the CL strategy.

## 4 The Role of the Antecedent

One of the questions we investigate in relation to the two innovative variants (NO and FP in BP) is what serves as a triggering factor for each one (Kato, 1993; Cyrino 1997; Creus, Menuzzi, 2004; Othero et. Al., 2018, Othero, Schwanke, 2018). Regarding the licensing of NOs in BP, it is already accepted that the language allows null objects whose antecedents are [-animate], specific, or not, as in (5a-b).

(5) a. [-ani; +spec]

*Comprei o casaco<sub>i</sub> sem experimentar \_\_\_\_<sub>i</sub>*  
 bought the coat without try Ø  
 'I bought the coat without trying.'

(Cyrino, 1993, p. 40)

b. [-ani, -spec]

*A instrutora visita escolas<sub>i</sub> antes de recomendar \_\_\_\_<sub>i</sub> aos alunos.*  
 The instructor.F visits schools before of recommend Ø to-the.pl students  
 'The instructor visits the schools before recommending them to the students.'

(Cyrino, 2017, p. 97, adapted)

It is also observed that, when animate, the antecedent is either specific (12a) or refers to a bare plural antecedent (6b).

(6) a. *Então os meus primos por parte de pai são todos muito*  
 then the.Pl my.Pl cousins for part of father are all very  
*mais velhos do que eu, eles têm idade, assim, de, eu chamo até de*  
 more old of-the that I they.M have age so of I call until of

*tio*                    *alguns, porque eu*                    *conheço*\_\_\_\_                    *desde*                    *pequeninha*.  
 Uncle                    some    because I                    know    Ø                    since    little

‘Then, my cousins by father are all very much older than me. Because of how old they are, I call some of them my uncle, also because I have known them since I was little.’

(Freire, 2000, *apud* Cyrino 2020, p. 392)

b. *Maria recolhe*                    *cachorros*<sub>i</sub>                    *por*                    *onde*                    *encontra*\_\_\_\_<sub>i</sub>. [+anim, -spec]  
 Maria collects                    dogs                    for                    where                    find                    Ø

‘Maria collects dogs wherever she finds them.’

(Creus and Menuzzi, 2004, p. 158, adapted)

When the antecedent expresses animacy and is specific, the anaphoric direct object must be realized, as shown by the contrast in (7):

(7) a. *Sabe*                    *a*                    *Maria*<sub>i</sub>? *Eu*                    *encontrei*                    *ela*<sub>i</sub>                    *ontem*                    *no*                    *cinema*  
 knows    the                    Maria    I                    found                    she                    yesterday                    in-the                    cinema

‘Do you remember Maria? I met her yesterday at the movies.’

(Creus and Menuzzi, 2004, p. 158, adapted)

b. \**A*                    *Maria*                    *encontrou*                    *o*                    *Pedro*<sub>i</sub>                    *mas*                    *não*                    *beijou* \_\_\_\_<sub>i</sub>.  
 the                    Maria                    found                    the                    Pedro                    but                    neg                    kissed                    Ø

‘Maria had a date with Pedro, but she didn’t kiss him.’

(Cyrino, 2020, p. 392, adapted)

When analyzing data from Afro-Brazilian Portuguese, Figueiredo (2004, 2009, 2019) finds variation between NOs and FPs in all contexts: [ $\pm$ animate,  $\pm$ specific], as shown in the examples (8) to (11). This resembles the popular urban Portuguese of Salvador, which suggests the behavior is shared by both BP varieties.

(8) [-anim, +spec]

*E*                    *esse*                    *dedo*                    *daqui*                    *mesmo*<sub>p</sub>                    *eu*                    *num...*                    *eu*                    *dobro*                    *ele*<sub>i</sub>.  
 and                    this                    finger                    of-here                    even                    I                    neg                    I                    flexes                    he

‘And this very finger, I can’t flex it.’

(9) [-anim, -spec]

*Nunca*                    *usei*                    *camisinha*<sub>i</sub>.                    *Elas*                    *sempre*                    *leva* \_\_\_\_<sub>i</sub>                    *eu*                    *dêxo* \_\_\_\_<sub>i</sub> *lá*  
 never                    used                    condom                    they.F                    always                    takes Ø I                    leave Ø                    there

*dentro*                    *guardarôpa*...

inside                    wardrobe

‘I never used condoms. They always bring them, but I leave them inside the wardrobe.’

- (10) [+ anim, + spec]  
**O pai dos meus filho<sub>i</sub>** é morto há quinze anos. **Mataro** \_\_\_\_\_  
 the father of-the.Pl my.Pl child is dead there-is fifteen years murdered Ø  
 ‘The father of my children is dead for fifteen years. He was murdered.’

- (11) [+ anim, - spec]  
**O nome certo dele** é xangó. ...**todo mundo vende ele** como pititinga  
 the name correct of-he is xangó every world sells he as pititinga  
 ‘The correct name of this fish is xangó, but everyone sells it as pititinga.’

(Figueiredo, Teixeira de Sousa, Quiroga, 2022, p. 78-79)

As pointed out in the latter section, referent’s animacy and specificity are fundamental characteristics to distinguish between the usage of NOs or FPs. Cyrino (2017) proposes that the variation between NOs and FPs could be related to the phenomenon known as *Differential Object Marking (DOM)*. DOM is common in Spanish, where [+animate, +specific] objects should be preceded by preposition *a*. Cyrino (2017) states that, in BP, direct objects could be subject to a differential marking due to the [+animate] feature. According to her, this feature would require the object to be an FP. However, we should consider that FPs do not categorically refer to animate antecedents, not even in the 19<sup>th</sup> century. Figueiredo (2019) finds two occurrences of [-animate] antecedents for FPs.

For Creus & Menuzzi (2004), animacy has a special role in the selection of NOs vs FPs given that this is the feature configuring basic generalizations of the system. Specificity, on the other hand, only seems relevant for the antecedents’ class. Thus, the authors propose that the relevance of animacy is related to the semantic gender (SG) distinction presented by the referents. This property would be crucial for the choice of the strategy, reducing the choice of NO/FP to a single opposition: the presence of SG in the antecedent. Specifically, the correlation established by the authors is that if the animate antecedent does not show semantic gender, then NO is used. In its turn, if the antecedent shows SG, then FP is used.

In our data, however, we observe FPs with antecedents that do not show SG, as seen below:

- (12) a. [+anim, -SG]  
*Num é porque eu não goste, é porque a pessoa que*  
 neg is why I neg like is why the person that  
**num é acostumada a ôvi, aí... pa num irritá ela,**  
 neg is used-to to hear then to neg irritate she  
*não toca música evangélica.*  
 neg play music gospel  
 ‘I don’t play gospel music, not because I don’t like it. It is to avoid irritating people who are not used to listening to gospel music.’

b. [+anim, -SG]

*Aquele p xe ali, a cesta sai a cinco reais. A gente trata ele*  
 that fish there the basket goes to five reais the people treat he  
*e faz os pacote de um real.*  
 and make the.pl packet of one real

‘The basket of that fish costs five reais. We clean the fish and separate them into packets that cost one real each.’

Nevertheless, as evidenced in the examples (8-12), popular Portuguese of Salvador and Afro-Brazilian Portuguese do not show categorical behavior with respect to animacy, specificity, and semantic gender. Variation is possible in every context, even though we believe these features act rather in terms of favoring one strategy instead of another, which should be evidenced in terms of frequency.

## 5 Popular Portuguese of Salvador

In this work, we investigate the linguistic features in the usage of FPs and NOs. We analyze colloquial Portuguese data sampled from three neighborhoods in Salvador: Itapu , Plataforma, and Liberdade.

The data was previously studied by Figueiredo, Teixeira de Sousa & Quiroga (2022) under a perspective of linguistic change. Therefore, the former analysis prioritized the age group variable, which, according to Weinreich, Labov & Herzog (1968), allows us to identify whether a phenomenon in variation is stable or undergoing change. The study also verified whether each neighborhood’s history could influence the choice of a particular variant. The authors’ findings are summarized in the Table 1.

**Table 1** – Data distribution per neighborhood<sup>2</sup>.

Variables	NO	FP	Total
Neighborhood	No./%	No./%	
Liberdade	359/91	32/9	391
Itapu�	528/86	87/14	615
Plataforma	463/83	96/17	559
Total	1350/86	215/14	1565

The Table 1 shows NO as the most frequent strategy in all three neighborhoods, suggesting a similar behavior to that pinpointed by sociolinguistic studies on the topic (Omena, 1978; Duarte, 1986; Figueiredo, 2004; among others). Although the percentages are close to each other, it is possible to verify a behavior that translates the history of their formation.

NOs are more frequent in Liberdade (91%) and, consequently, FPs are less frequent there (9%). This result is expected, since, according to Ramos (2007), Liberdade is considered

<sup>2</sup> Figueiredo, Teixeira de Sousa, Quiroga (2022, p. 83).



a national territory of Afro-Brazilian culture by the Brazilian Ministry of Culture. According to Figueiredo, Teixeira de Sousa & Quiroga (2022), the prevalence of NOs reveals that conservative linguistic usage of Afro-Brazilian Portuguese is thereby preserved.

Itapuã displays an intermediate behavior, with 86% of NO and 14% of FP. Figueiredo, Teixeira de Sousa & Quiroga (2022) note that such behavior seems to reflect the neighborhood's history. Settlement in the region spiked due to the development of agricultural activities in the 17<sup>th</sup> century. These were suburban *quilombos* (Afro-Brazilian communities), and their current traits became prevalent upon the settlement of fishermen villages in the area and its further transformation into a holiday destination. Thus, the neighborhood now consists of people from distinct social classes and educational levels; it is heavily populated, and its beaches and shopping amenities attract people from different places. Such extralinguistic factors allow the authors to consider that not only was linguistic diversity the norm in the past, but it continues to be in the present.

Finally, Plataforma is the neighborhood with the lowest percentage of NO occurrence (83%) — and, consequently, with the highest percentage of FPs (17%). Plataforma is a heavily populated, working-class neighborhood. It has its origins during the implementation of the railroad system, completed in 1860. In addition, the neighborhood was also home to industrial workers — manufacturing was an emerging economic activity in the proximities. The region, therefore, lacks an identity. Figueiredo, Teixeira de Sousa & Quiroga (2022) state that it is a lower income neighborhood, it does not bear an ethnic identity, as observed in Liberdade, or a touristic one, as in Itapuã. This could be an important factor in favor of the more stigmatized variant, FP.

For this analysis, we used the corpus built by professors Dante Lucchesi and Gredson Santos for the Vertentes Project in 2008.<sup>3</sup> The corpus was set up within the Variationist Sociolinguistics framework (Labov, 1983). The study analyzed the speech of 36 informants — 12 from each neighborhood. The samples were obtained through interviews with a minimum of 40 minutes. The speakers were divided into three age groups: Group 1, 25-35 years old; group 2, 45-55 years old; and group 3, 60+ years old.

## 6 Data Analysis<sup>4</sup>

The observations from the corpus were classified with respect to the following linguistic independent variables: animacy, semantic gender, and specificity. The extralinguistic independent variables are age group, educational level, and neighborhood.

In this section, we firstly conduct an exploratory analysis, evaluating the association between the independent variables with our dependent variable: The anaphoric strategy, whether FP or NO. Exploratory analyses are useful to determine the significance of each variable with respect to our main questions and to aid us in the interpretation of the data.

---

<sup>3</sup> These corpora were recorded before the Ethics Committee for Research with Human Subjects regulated this kind of sample.

<sup>4</sup> Taking into account that abstract features do not license [+animate] antecedents, we decide to analyse abstract data separately.

After this exploratory stage, we use Logistic Regression analysis to establish to which extent each variable might affect the choice of FP vs NO as an anaphoric strategy. Finally, we discuss the analysis, emphasizing the relationships between the results and the socio-history of the BP variants we observed.

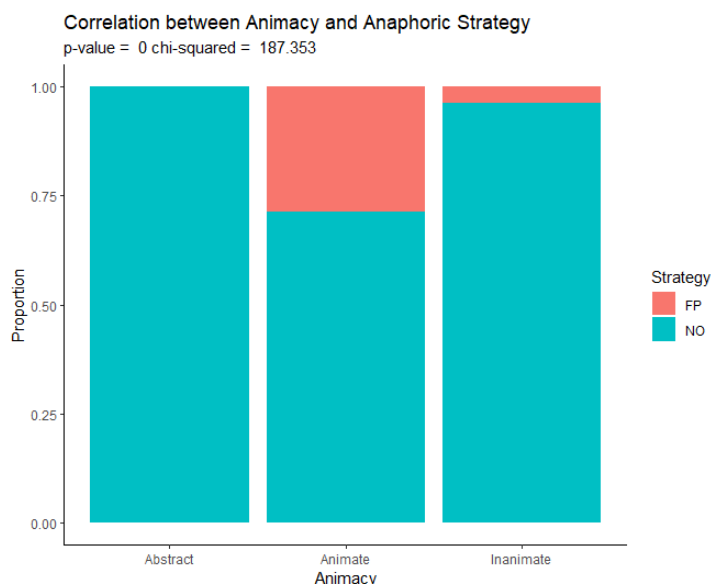
## 6.1 Exploratory Analysis

Firstly, we analyzed data distribution in relation to animacy. As explained before, literature has often associated animacy with a higher occurrence of FPs vs NOs, therefore it should be considered an important explanatory variable. The Table 2 shows the distribution of each anaphoric strategy in relation to animacy:

**Table 2** – Data distribution with respect to animacy.

	<b>Abstract</b>	<b>Animate</b>	<b>Inanimate</b>
Full Pronoun (FP)	0 (0%)	179 (28,6%)	25 (3,2%)
Null Object (NO)	133 (100%)	445 (71,4%)	640 (96,8%)
Totals	133	624	665

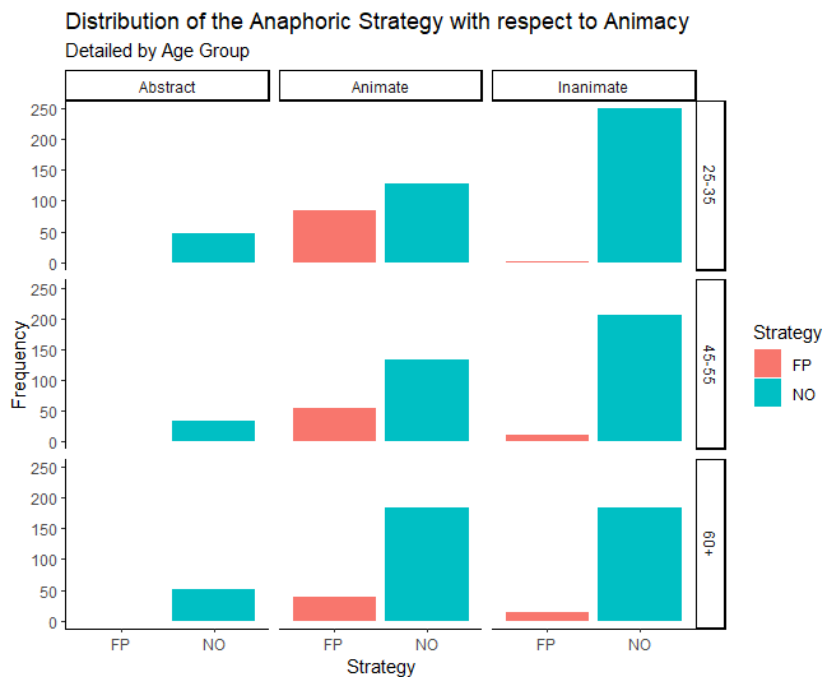
A *chi-squared* test was conducted to check whether the relationship between both variables is statistically significant, and it allowed us to reject the null hypothesis of independence of the variables with a near zero *p-value* ( $p=2.07 \times 10^{-41}$ ,  $\chi^2 = 187.353$ ). Below is a plot of the distribution (Figure 1).



**Figure 1** – Correlation between Animacy and Anaphoric Strategy.

We can verify that NOs are always the preferred strategy. However, there is a clear relationship between the strategies with animacy. Despite not being the preferred strategy, FPs tend to be chosen more frequently with animate referents than with abstract or [-animate] ones.

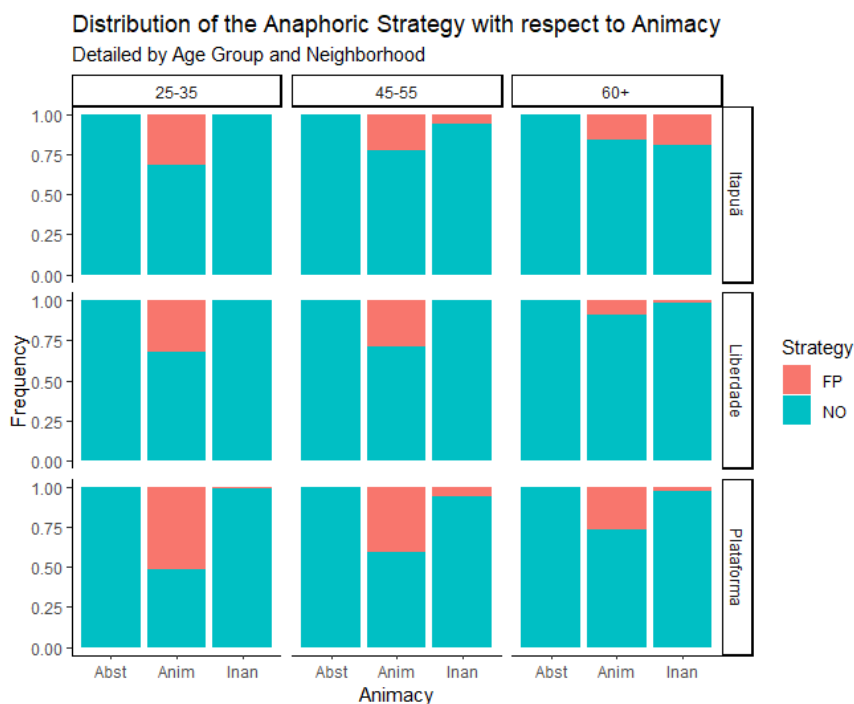
It is interesting to see how this distribution relates to the age groups. The visualization below splits the data according to the three age group tiers (Figure 2).



**Figure 2** – Distribution of the Anaphoric Strategy with respect to Animacy detailed by Age Group.

We can clearly see that speakers over 60 years old tend to have a generally lower production of FPs and to favor NOs as their anaphoric strategy. The youngest age group, on the other hand, has a higher production of FPs, mainly with animate precedents.

It is also remarkable that the neighborhoods differ with relation to the distribution of animacy for FPs among different age groups. The plot below shows the proportion of each strategy with abstract (Abst), animate (Anim) and [-animate] (Inan) referents in each age group and each neighborhood (Figure 3).



**Figure 3** – Distribution of the Anaphoric Strategy with respect to Animacy detailed by Age Group and Neighborhood.

Significant differences can be observed within the 60+ age group across different neighborhoods. One can see that the elderly from Plataforma already show the same tendencies of favoring FPs for animate referents as seen in the younger group in all neighborhoods. It is just the proportion of FPs vs NOs that is augmented with animate referents in the younger generations. Reversely, in Itapuã and Liberdade, the elderly show a more even distribution of the strategies, regardless of the referent. However, FPs are more common in Itapuã. Our *chi-squared* tests with the 60+ age group for each neighborhood corroborate that, in Plataforma, FPs relies on animacy to occur ( $p = .00049$ ,  $\chi^2 = 12.147$ ). In contrast, Liberdade ( $p = .1623$ ,  $\chi^2 = 1.9526$ ) shows independence between anaphoric strategy and animacy. Itapuã, stands in-between though, with a slight difference in the counts of NOs and FPs for animate and inanimate antecedents. Such difference is enough to be considered statistically significant ( $p = 0.0371$ ,  $\chi^2 = 4.3502$ ), suggesting dependence between the anaphoric strategy and animacy. Nevertheless, the p-value obtained from the test with data from Itapuã is way higher than the p-value obtained from Plataforma.

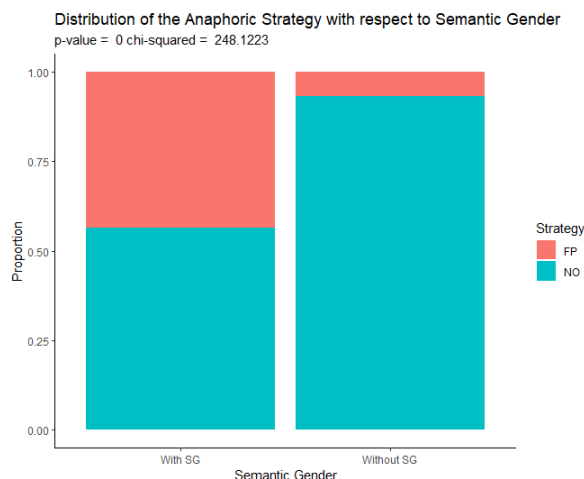
In summary, observations of the older age group show no relation between animacy and the choice of the anaphoric strategy in Liberdade. This relationship starts to appear in younger age groups. Plataforma, on the other hand, shows this relationship across all age groups. Itapuã shows a way more even distribution than in Plataforma, however with a statistically significant difference in the frequency of PFs for animate and for inanimate antecedents. It is interesting to recall that this description matches the neighborhoods' historical constitution: Liberdade and Itapuã originated from Afro-Brazilian communities (Liberdade to a much greater extent), which are associated with a systematic production of NOs. Plataforma, on the other hand, is inhabited by people from more diverse backgrounds and thus a higher production of FPs from the elderly is expected.

We also analyzed the data with respect to whether the antecedents bear semantic gender or not. According to Creus and Menuzzi (2004), rather than animacy itself, it is the presence of an antecedent with semantic gender that triggers the occurrence of FPs instead of NOs. Semantic gender is distributed in our data according to the Table 3.

**Table 3** – Data distribution with respect to the presence of Semantic Gender

	With SG	Without SG
Full Pronoun (FP)	126 (43%)	78 (7%)
Null Object (NO)	164 (57%)	1054 (93%)
Totals	290	1132

As with Animacy, a *chi-squared* test will suggest dependency between semantic gender and the choice of the anaphoric strategy ( $p = 6.66 \times 10^{-56}$ ,  $\chi^2 = 248.1223$ ). A visualization of the data above is shown in Figure 4.



**Figure 4** – Correlation between the Anaphoric Strategy and the presence of Semantic Gender

As with animacy, though, NOs comprise most of the observations even when the antecedents bear semantic gender. FPs are more constrained when their antecedents do not have semantic gender and tend to happen mostly with antecedents having semantic gender. This behavior matches that of animacy.

When we split the data by neighborhood and age group, however, we see that the distribution follows a different pattern than that of animacy. Here is the plot (Figure 5).

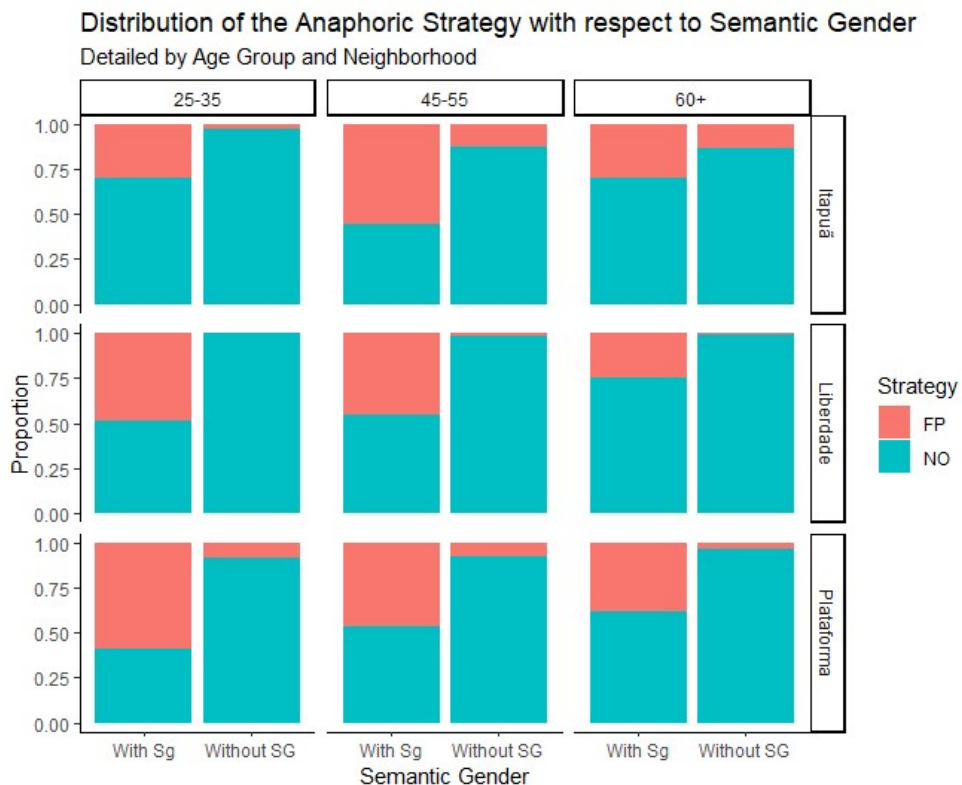
The 60+ group's distribution of FPs is slightly more even in Itapuã, but uneven in Liberdade and Plataforma. This contrasts with what we observed for animacy, where FPs were evenly distributed in both animate and [-animate] contexts among the elderly of both Itapuã and Liberdade. Plataforma keeps the same status across all ages, and FPs are systematically restricted to occur referring to antecedents with semantic gender.

It is important to notice that semantic gender and animacy are themselves extremely correlated, as show in Figure 6.

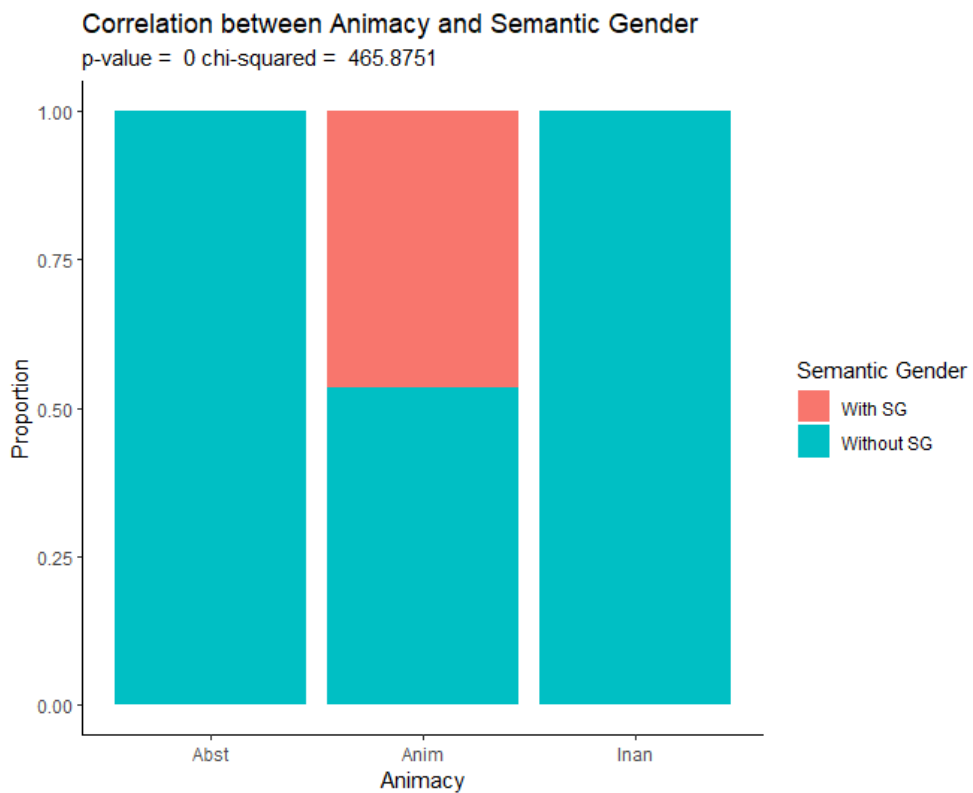
Nevertheless, this correlation varies depending on the age group and neighborhood. We can see below that the correlation stands across all ages in Plataforma, and everywhere among the youngest. Elder and adult speakers from Itapuã seldom produced data without semantic gender (Figure 7).

A final linguistic variable observed is Specificity. It refers to the definiteness and specificity of the referent. The variable is encoded in five levels, being: Definite specific, definite non-specific, indefinite specific, indefinite non-specific, and generic. Below is the distribution of the variable with respect to the anaphoric strategy (Table 4), and the visualization (Figure 8).

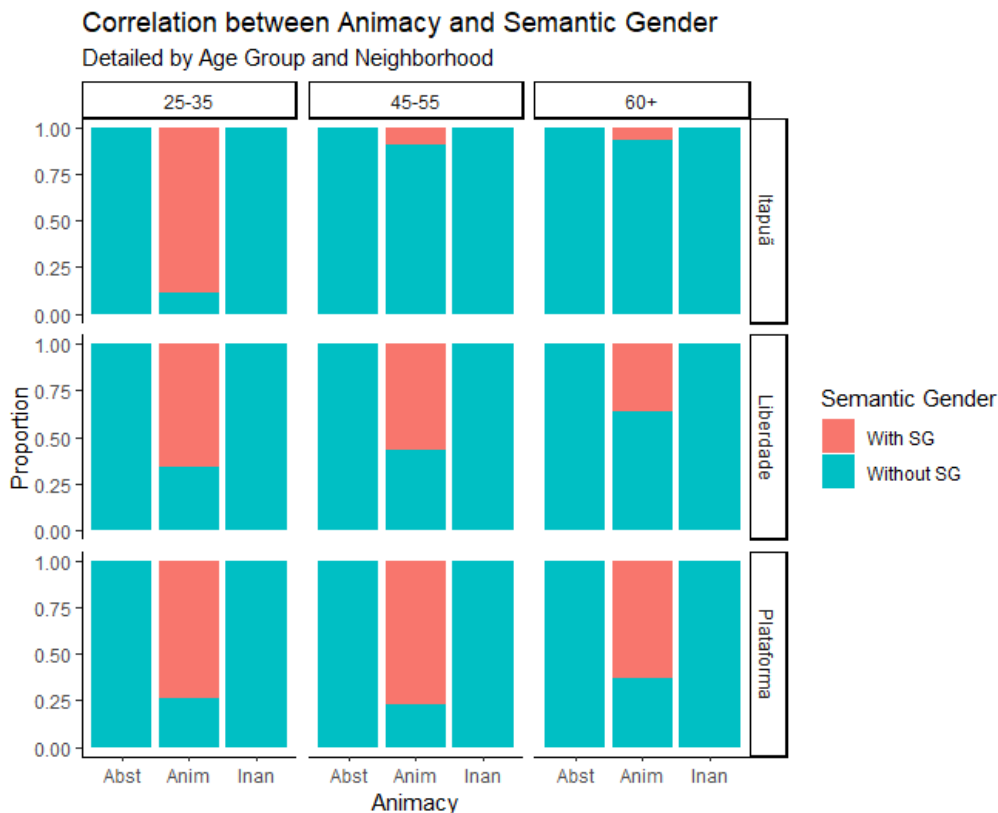
We see that FPs are preferred with definite referentiality rather than with indefinite referentiality. Although this seems to be relevant information, it is important to understand the extent to which referentiality correlates with the other linguistic variables. We have used a multivariate method called Correspondence Analysis (CA) to compare the three variables. CA is an extension of Principal Component Analysis (PCA) to categorical data, meaning that, through this method, we can plot categorical variables in a two-dimensional space and see whether their values cluster together or not. In CA, when two variables are given, values that frequently happen together will be plotted closer together in the biplot. In its turn, values that seldom happen together will be distant from one another.



**Figure 5** – Distribution of the Anaphoric Strategy with respect to Semantic Gender detailed by Age Group and Neighborhood.



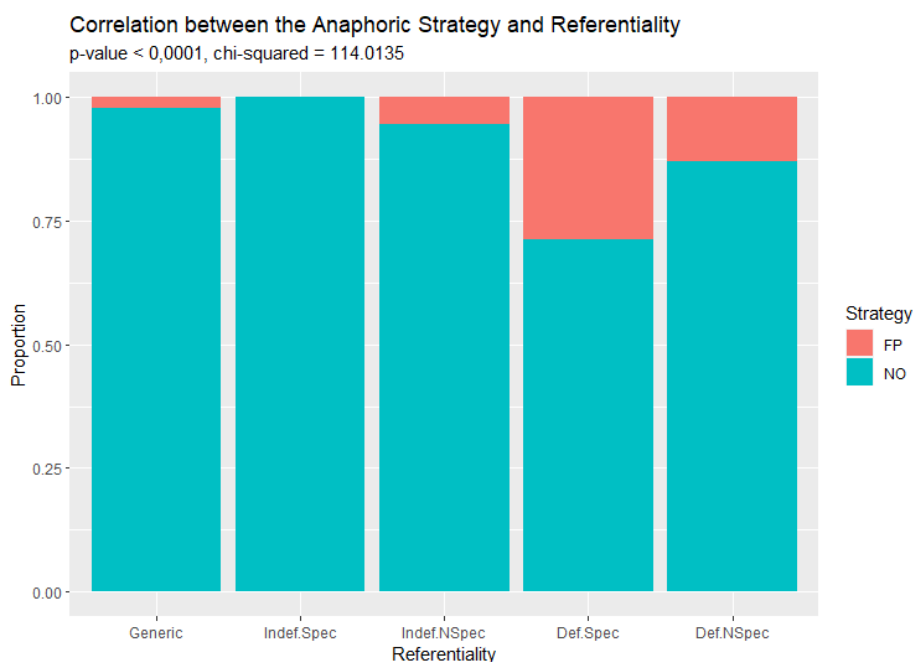
**Figure 6** – Correlation between Animacy and Semantic Gender.



**Figure 7** – Distribution of Animacy and Semantic Gender detailed by age group and neighborhood.

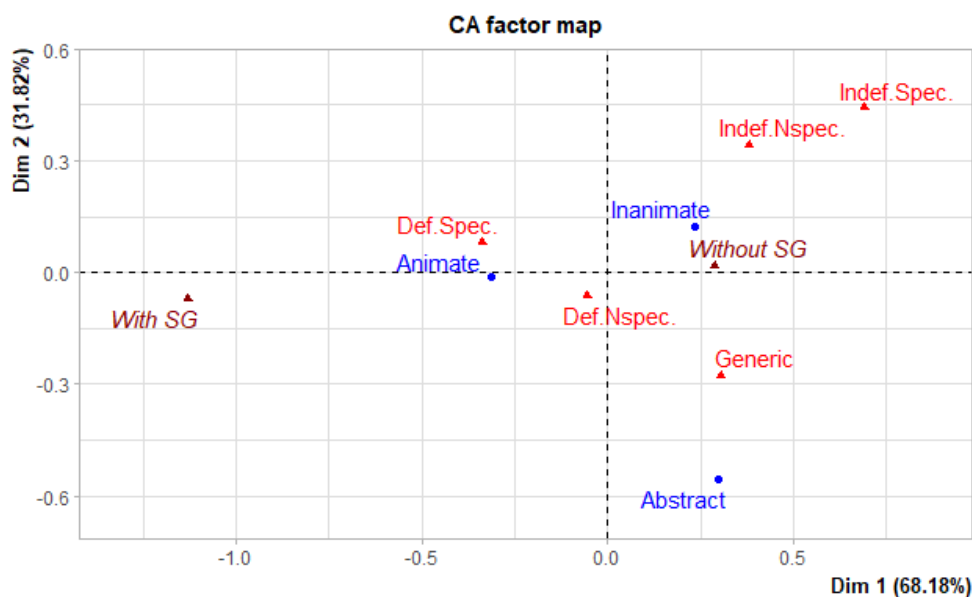
**Table 4** – Distribution of the Anaphoric Strategy with respect to Referentiality

	Def. Spec.	Def. Non-Spec.	Indef. Spec.	Indef. Non-Spec.	Generic
Full Pronoun (FP)	116	72	0	10	6
Null Object (NO)	287	479	24	174	254



**Figure 8** – Correlation between Anaphoric Strategy and Referentiality.

We used the R language's FactoMineR library, which provides a well-optimized CA algorithm that outputs a biplot for a given contingency table. We can see the result in Figure 9.



**Figure 9** – Correspondence Analysis considering Animacy and Referentiality.

Correspondence Analysis relates two main variables, which must have three or more levels. Our chosen main variables are Animacy and Referentiality. We also added Semantic Gender as a supplementary variable. Supplementary variables help us to interpret the plot. What we see above is that the horizontal dimension is related to the semantic gender, with the presence of SG on the left side and its absence on the right. Animate and Def. Spec. observations are very close to each other and are the closest to “With SG”. Most of the values are clustered around “Without SG”.

The first interesting aspect of the plot is that the values for Animacy and Referentiality tend to cluster together: Abstract with Generic, Animate with Def. Spec. and Def. NSpec., Inanimate with Indef. Spec. and Indef. NSpec. Conversely, a *chi-squared* test between both variables yields an extremely low p-value ( $p= 9.23 \times 10^{-31}$ ,  $\chi^2=161.14$ ).

Secondly, when we add anaphoric strategy as a supplementary variable, we begin to understand the distribution of the data (Figure 10).

Here, the different values for independent variables seem to orbit around the values for the anaphoric strategy (NO and FP). Most of the values are close to NO, which explains its general prevalence. However, “With SG” is the most distant from NO, which suggests that observation with this value will be extremely likely to be associated with a FP. It is also curious to see that “animate” (and Def. Spec.) are located in between both strategies. This suggests that they can be equally associated with FPs and NOs. In any case, referents that are definite specific, animate and bear semantic gender are the only ones with enough likelihood to be referenced by an FP.

We saw earlier that neighborhood and age group are also important factors to be considered with respect to the distribution of NOs and FPs. We conducted another CA to include these variables (Figure 11).



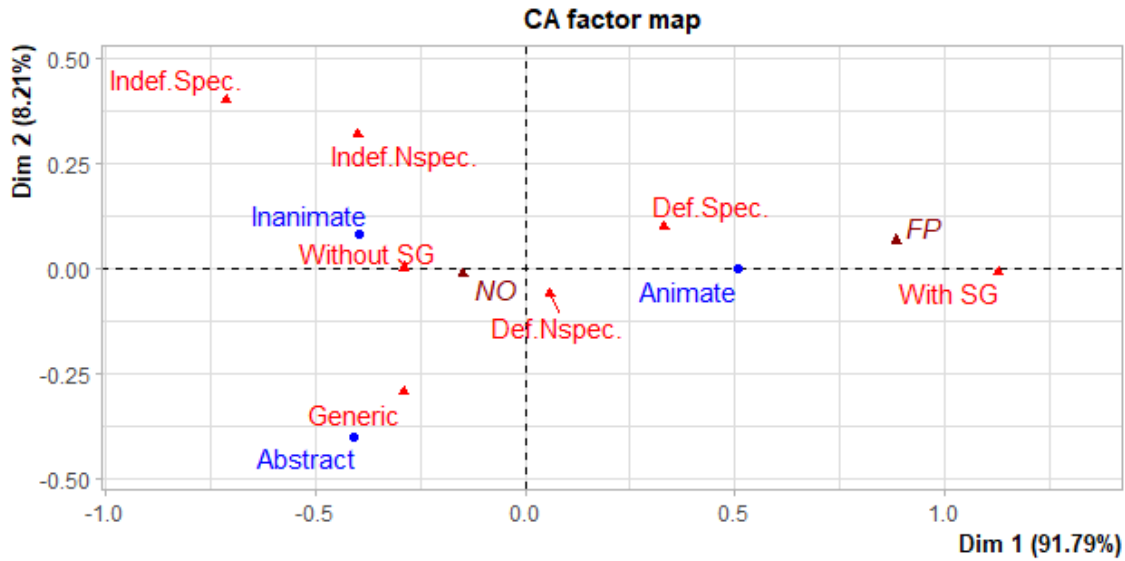


Figure 10 – Correspondence Analysis considering the anaphoric strategy and the other linguistic variables.

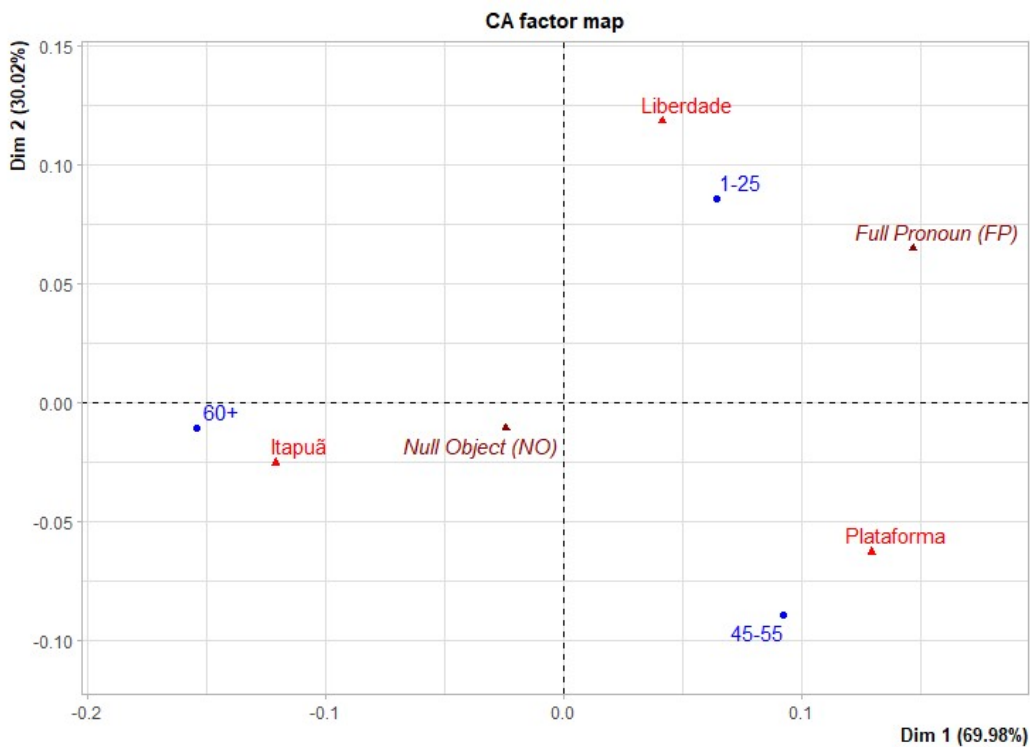


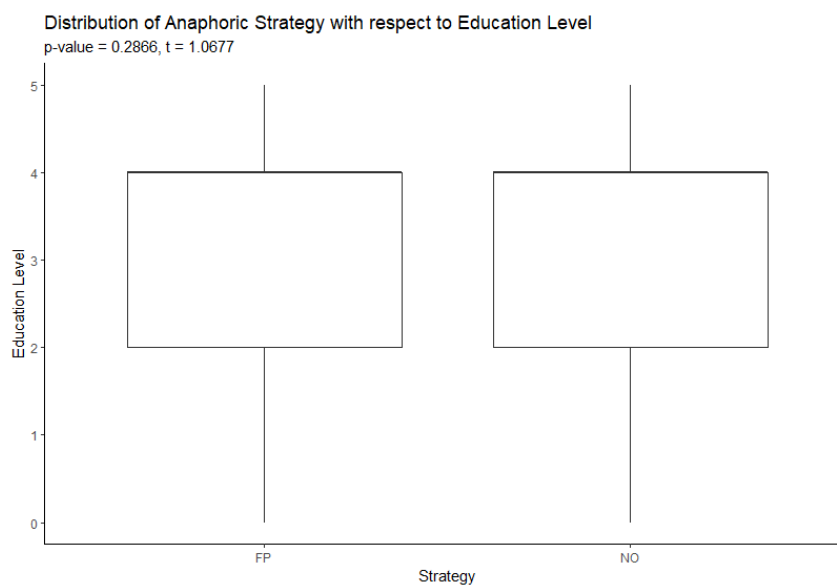
Figure 11 – Correspondence Analysis considering Age Group and Neighborhood.

As with the linguistic variables, we see that most values orbit around the NO anaphoric strategy with FP being closer to the younger age group. NOs are also associated with the elderly. With respect to the neighborhoods, FPs are mostly likely to happen in Plataforma and Liberdade whereas Itapuã would prefer NOs over FPs even in linguistic contexts favoring the

latter. What gives Itapuã this characteristic is the fact it is associated with the 60+ age group. Liberdade has a younger population. Even if the elders in that neighborhood would avoid FPs, the region has a higher number of young people. Plataforma's situation is stable, sitting in between NO and FP and closely related to the adult population.

CA is an important tool to verify whether two categorical variables are highly correlated in any context. We saw that animacy and referentiality are systematically clustered together as well as age group and neighborhood. This may have a great impact when analyzing data through regression models and, most of the time, it is better to choose one variable instead of the other. While we are not sure which is the most important, age group or neighborhood, we preferred to exclude referentiality from our logistic regression analysis. Referentiality is a multi-level variable and its interpretation can be difficult. Animacy, on the other hand, has been widely studied across literature and has only three levels, which makes the regression output easier to interpret.

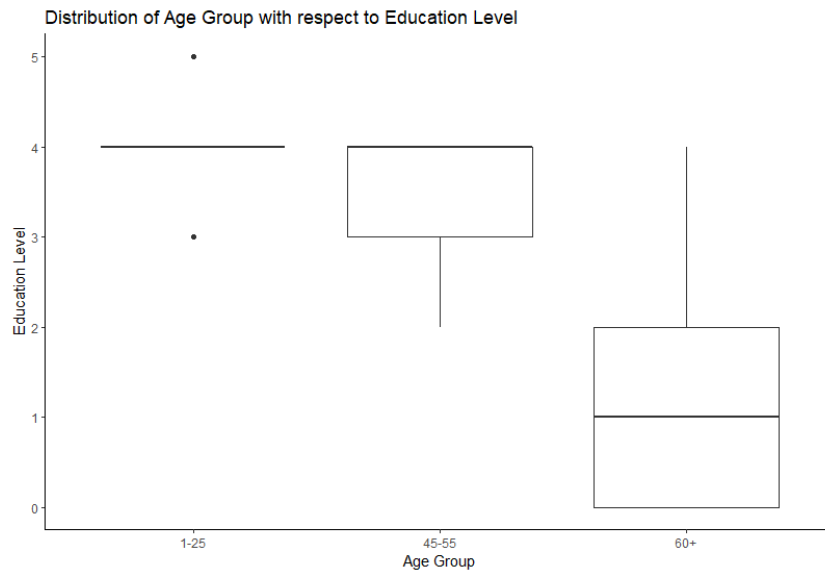
Educational level is a social variable which has not been considered yet. This variable was classified into six different levels — 0 for illiterate individuals, and 1-5 indicates the highest grade attended by the subject: 1 -1<sup>st</sup> grade; 2 - 2<sup>nd</sup> grade; 3 - 3<sup>rd</sup> grade; 4 -4<sup>th</sup> grade, 5 -5<sup>th</sup> grade or above. This is a numeric variable, to be read as a scale. Below is the boxplot with respect to each anaphoric strategy (Figure 12).



**Figure 12** – Distribution of the Anaphoric Strategy with respect to Educational Level.

As we can see, educational level seems to be evenly distributed between the observations of both strategies. Both have a median at 4 and the difference in means is insignificant as suggested by the *t-test* ( $p = .2866$ ,  $t = 1.0677$ ).

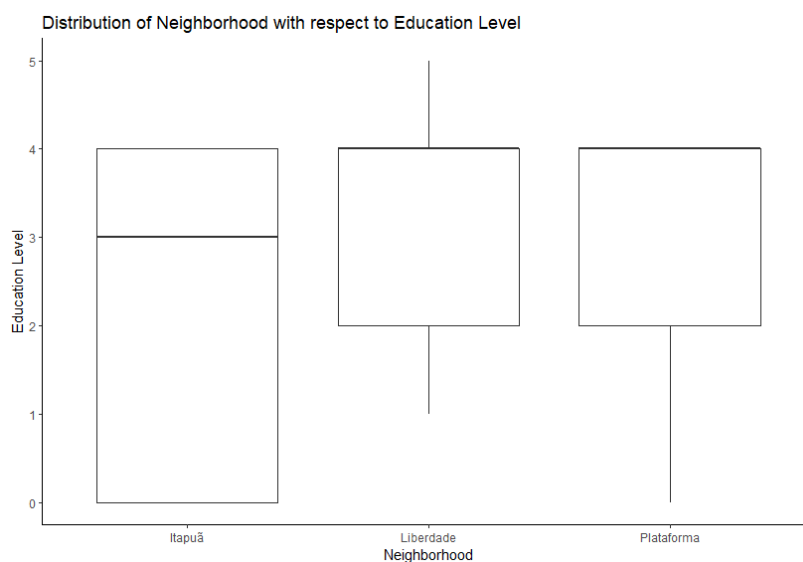
This variable is interesting in association with the age group. Below we see the distribution of educational level per age group (Figure 13).



**Figure 13** – Distribution of Educational Level per Age Group.

We see a striking difference between the 60+ group and other age groups. The former's median is 1 while the latter is 4. In addition, the youngsters are more homogeneous in this respect than other generations, being the 60+ the most heterogeneous group.

The distribution of educational level among neighborhoods is somewhat more homogeneous, with Itapuã being the neighborhood with the lowest median (3). Plataforma and Liberdade have their median at 4. Below, in Figure 14, is the boxplot.



**Figure 14** – Distribution of Education with respect to Neighborhood.

## 6.2 Logistic Regression

We used Logistic Regression to establish to which extent each variable cooperates in the selection of either FP or NO as the anaphoric strategy. We evaluated animacy, specificity, semantic gender, neighborhood, educational level, and age group variables, these variables gave us the highest concordance rate ( $C = 0.854$ ) with no detectable multicollinearity (variance inflation factors  $< 5$ )

Below are the results for the logistic regression. We present the output of the coefficients of the model as given by the R *rms* package's logistic regression function (*lrm*).<sup>5</sup> These coefficients correspond to the logarithm of the odds in favor of NO. Therefore, positive coefficients favor NO and negative coefficients favor FP. We also present the Odds Ratio (in favor of NOs) and Relative Weight for the corresponding coefficients, since the reader may have different preferences on which form of data to interpret. The standard error and p-value for each coefficient is presented, followed by a column called significance, relative to the level of significance in favor of each strategy. For instance, “+++ NO” means there is a p-value of less than .001 in favor of NO. “++ FP” means a p-value of less than .01 in favor of FP. “+ FP” means a p-value of less than .05 in favor of FP. Finally, “. NO” means that there is a p-value less than .1 in favor of NO (Table 5).

**Table 5** – Logistic Regression Output.

Variable	Coefficient	Std.Error	Odds Ratio (NO)	Relative Weight	p-value	Significance
(Intercept)	12.6931	33.7774	>1000	0.999	0.7071	
Animate	-10.0437	33.7737	<0.001	<0.001	0.7662	
Inanimate	-8.4528	33.7741	<0.001	<0.001	0.8024	
Without SG	1.0939	0.2802	2.916	0.749	<0.0001	+++NO
Indef. Spec.	7.9321	76.9257	2705	0.999	0.9179	
Indef. Nspec.	-1.1490	0.5475	0.319	0.240	0.0358	+FP
Definite Spec.	-2.0436	0.4625	0.129	0.114	<0.0001	+++FP
Definite Nspec.	-2.2893	0.4874	0.099	0.092	<0.0001	+++FP
Liberdade	-0.1106	0.2971	0.900	0.472	0.7096	
Plataforma	-0.9053	0.2430	0.403	0.287	0.0002	+++FP

The model shows that antecedents without semantic gender favor NO significantly (2.916 times more likely). FPs are, on the other hand, significantly favored with most antecedents with respect to specificity, except Indefinite Specific ones. By inverting the Odds Ratio (1/Odds Ratio), we have that Indefinite Non-Specific antecedents are 3.13 times more likely to be referred by FPs, Definite Specific antecedents are 7.71 times and Definite Non-Specific are 10.06 times. The neighborhood of Plataforma is also a significant factor for the occurrence of FPs instead of NOs, where FPs are 2.47 more likely to occur.

<sup>5</sup> The formula used is  $\text{variable} \sim \text{animacy} + \text{specificity} + \text{semantic\_fender} + \text{neighborhood}$ .

## 7 Final remarks

In this paper, we attempted to describe the distribution between FP and NO in relation to their antecedent by analyzing data collected in three neighborhoods from Salvador, Bahia, Brazil. Considering the origins of NO and nominative full pronouns in BP, we hypothesized that the frequency of the FP and their antecedent as [+animate] would be higher in the youngest group.

The logistic regression model showed us that definite and indefinite non-specific antecedents and living in Plataforma are the main factors in favor of the occurrence of FPs. On the other hand, NOs are significantly favored when the antecedent doesn't have semantic gender. It is worth mentioning that the model did not confirm age group as a relevant variable. This is probably due to the high correlation observed between the age group and neighborhood variables.

The model matches what we have been observing in the socio-history of the BP of Salvador, which is largely influenced by Afro-Brazilian Portuguese. Plataforma is the only neighborhood where FPs are observed across all age groups. Nevertheless, in Plataforma, antecedents with semantic gender will favor NOs. This tells us that not only is Plataforma the neighborhood where FPs are more prevalent, but also that it is where animacy/semantic gender is a criterion for the distribution of the different anaphoric strategies.

With respect to the linguistic variables, antecedents without semantic gender are more relevant in favoring NOs than antecedents with semantic gender in favoring FPs. Definiteness (and non-specific indefiniteness) tends to be the relevant criterion for FPs favoring.

## REFERENCES

- ARRUDA, N. C. *A realização do objeto direto no português brasileiro culto falado: um estudo sincrônico*. Dissertação (Mestrado) – Universidade Estadual Paulista, 2006.
- AVERBUG, M. C. G. *Objeto direto anafórico e sujeito pronominal na escrita de estudantes*. Dissertação (Mestrado) – Universidade Federal do Rio de Janeiro, 2000.
- CREUS, S; MENUZZI, S. O papel do gênero na alternância entre objeto nulo e pronome pleno em português brasileiro. *Revista da ABRALIN*, n. 3, p. 149-176, 2004. <https://doi.org/1.5380/rabl.v3i1/2.52670>
- CYRINO, S. M. L. Null objects in Brazilian Portuguese. *Cuadernos de La ALFAL*, v. 12, p. 387-341, 2020.
- CYRINO, S. M. L. Reflexões sobre a marcação morfológica do objeto direto por a em português brasileiro. *Estudos Linguísticos e Literários*, n. 58, p. 83-103, 2017.
- CYRINO, S. M. L. *O objeto nulo no português do Brasil: um estudo sintático-diacrônico*. Londrina: UEL, 1997.
- CYRINO, S. M. L. Observações sobre a mudança diacrônica no Português do Brasil: objetos nulo e clíticos. In: ROBERTS, I.; KATO, M. (ed.). *Português Brasileiro: uma viagem diacrônica*. 2. ed. Campinas: Editora da UNICAMP, 1993.
- CYRINO, S. M. L., DUARTE, M. E. L., KATO, M. A. Visible subjects and invisible clitics in Brazilian Portuguese. In: KATO, M. A.; NEGRÃO, E. V. (ed.). *Brazilian Portuguese and the Null Subject Parameter*. Frankfurt ; Madri: Vervuert; Iberoamericana, 2020. p. 55-73

- DUARTE, M. E. L. *Varição e Sintaxe: clítico acusativo, pronome lexical e categoria vazia no português do Brasil*. Dissertação (Mestrado) – Pontifícia Universidade Católica de São Paulo, 1986.
- FIGUEIREDO, C. O objeto direto anafórico. In: GALVES, C.; LOBO, T. (org.). *O português escrito por afro-brasileiros no século XIX: as atas da Sociedade Protetora do Desvalidos*. Salvador: Edufba, 2019. v. 1. p. 175-216.
- FIGUEIREDO, C. Objeto nulo versus estratégias pronominais no português rural do estado da Bahia e no de Portugal. *Revista Confluência*, Rio de Janeiro, v. 46, p. 79-114, 2014.
- FIGUEIREDO, C. *O objeto nulo no português rural baiano*. Teoria Temática e Elipse de DP. Tese (Doutorado) – Universidade Federal da Bahia, 2009.
- FIGUEIREDO, C. O objeto direto anafórico no dialeto rural afro-brasileiro. *Revista Estudos Linguísticos e Literários*, Salvador, n. 33/34, p. 45-68, 2006.
- FIGUEIREDO, C. *O objeto direto anafórico no dialeto rural afro-brasileiro*. Dissertação (Mestrado) – Universidade Federal da Bahia, 2004.
- FIGUEIREDO, C., TEIXEIRA DE SOUSA, L., QUIROGA, A. O objeto direto anafórico no português popular de Salvador: objeto nulo e pronome pleno. In: BARROS, I.; CARNEIRO, J.; PARANHOS, R. A.; COELHO, T. (ed.). *Português Baiano: de norte a sul, de leste a oeste*. Salvador: EDUFBA, 2022. p. 71-90.
- GALVES, C. L'objet nul et la structure de la proposition en portugais du Brésil. *Revue des Langues Romanes*, v. 93, n. 2, p. 305-336, 1989.
- GALVES, C. Pronomes e categorias vazias em português do Brasil. *Cadernos de Estudos Linguísticos*, v. 7, p. 107-306, 1984. <https://doi.org/1.20396/cel.v7i.8636687>
- HUANG, C.T.J. On the distribution and reference of empty pronouns. *Linguistic Inquiry*, v. 15, p. 531-574, 1984.
- KATO, M. A. Null objects and VP-ellipsis in European and Brazilian Portuguese. In: QUER, J. et al. (ed.). *Romance languages and linguistic theory*. Amsterdam: John Benjamins, 2001. p.131-154. <https://doi.org/1.1075/cilt.245.09kat>.
- KATO, M. A. The distribution of null and pronominal objects in Brazilian Portuguese. In: *Linguistic perspectives on the romance languages: selected papers from the XXI linguistic symposium on romance languages*. Amsterdam: John Benjamins. 1993. p. 225-235.
- LABOV, W. *Modelos Sociolinguísticos*. Madrid: Cátedra, 1983.
- MACEDO-COSTA, T. *Um estudo diacrônico das variadas realizações do objeto direto anafórico na imprensa baiana dos séculos XIX e XX*. Dissertação (Mestrado) – Universidade Estadual de Campinas, 2012.
- OMENA, N. P. *Pronome pessoal de terceira pessoa: suas formas variantes em função acusativa*. Dissertação (Mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, 1978.
- OTHERO, G. A; SCHWANKE, C. Retomadas anafóricas de objeto direto em português brasileiro escrito. *Revista de Estudos da Linguagem*, v. 26, n. 1, p. 147-185, 2018. <https://doi.org/1.17851/2237-2083.26.1.147-185>
- OTHERO, G. A., CYRINO, S. M. L., ALVES, L. T. M., ROSITO, R. B. V., SCHABACH, G. R. Objeto nulo e pronome pleno na retomada anafórica em PB: uma análise em corpora escritos com características de fala. *Revista da Anpoll*, v. 1, n. 45, p. 68-89, 2018. <https://doi.org/10.18309/anp.v1i45.1113>

RAMOS, M. E. R. *Território e afrodescendente: leitura de cidade através do bairro da Liberdade*. Dissertação (Mestrado) – Universidade Federal da Bahia, 2007.

RAPOSO, E. On the null object in European Portuguese. In: JAEGGLI, O.; SILVA-CORVALAN, C. (ed.). *Studies in Romance Linguistics*. Dordrecht: Foris Publications, 1986.

SCHWENTER, S. A. Null objects across South America. In: *Selected proceedings of the 8<sup>th</sup> Hispanic Linguistics Symposium*. Somerville: Cascadilla Press, 2006.

SCHWENTER, S. A. Two kinds of differential object marking in Portuguese and Spanish. In: AMARAL, P.; CARVALHO, A. M. (org.) *Portuguese-Spanish Interfaces: Diachrony, synchrony, and contact*. Philadelphia: John Benjamins Publishing, 2014. p. 237-260.

TEIXEIRA DE SOUSA, L.; FIGUEIREDO, C. O pronome pleno no português afro-brasileiro em construções de ODA: Um possível percurso da mudança. *Cadernos de Estudos Linguísticos*, v. 64, p. 1-15, 2022.

WEINREICH, U; LABOV, W; HERZOG, M. I. *Fundamentos Empíricos para uma Teoria da Mudança Linguística* São Paulo: Parábola, 2006 [1968].