Null subjects and null objects in Brazilian Portuguese: correlations and change

Sujeitos e objetos nulos em português brasileiro: correlações e mudança

Gabriel de Ávila Othero

Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul / Brasil

gabriel.othero@ufrgs.br

http://orcid.org/0000-0002-2060-6312

Melissa Lazzari

Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul / Brasil

melissaglazzari@gmail.com

http://orcid.org/0000-0002-7819-9326

Abstract: Pronominal and null subjects and anaphoric direct objects are amongst the main grammatical phenomena that differentiate Brazilian and European varieties of Portuguese (cf. CYRINO; MATOS, 2016; DUARTE; FIGUEIREDO SILVA, 2016). In Brazilian Portuguese (BP), clauses with overt referential pronominal subjects are preferred over clauses with null subjects – which are restricted to some contexts (cf. AYRES, 2021, DUARTE; REIS, 2018); on the other hand, clauses with null anaphoric direct objects, especially 3rd person direct objects, are preferred over clauses with pronominal direct object (cf. CYRINO, 1993, 1997). The literature relating both phenomena in the history of BP shows that there has been a change in the following direction: null objects and overt pronominal subjects have had an increased frequency throughout the last centuries. Here we argue for a relation connecting both phenomena in BP: both are sensitive to the semantic gender of the referent when it comes to 3rd person pronoun realization. We support our hypothesis by bringing data from our previous work and work from colleagues. We also argue for a crucial difference between both phenomena: even though null objects are stabilized in BP grammar, we present empirical

data that show evidence for change in apparent time in 3rd person pronominal subjects. We have collected and analyzed data from a contemporary oral corpus (LínguaPOA) and here we show that (i) the asymmetry between 1st and 2nd person subjects on the one hand and 3rd person subjects on the other no longer exists; (ii) 1st and 2nd person null subjects are a stabilized grammatical phenomenon; and (iii) 3rd person null subjects are still changing.

Keywords: change; null subject; pronominal subject; anaphoric direct object; Brazilian Portuguese.

Resumo: Sujeitos pronominais e nulos e objetos diretos anafóricos estão entre os principais fenômenos gramaticais que diferenciam as variedades do português brasileiro e europeu (cf. CYRINO; MATOS, 2016; DUARTE; FIGUEIREDO SILVA, 2016). No português brasileiro (PB), as orações com sujeitos pronominais referenciais são preferidas em detrimento das orações com sujeitos nulos – que são restritas a alguns contextos (cf. AYRES, 2021, DUARTE; REIS, 2018); por outro lado, orações com objetos diretos anafóricos nulos, especialmente objetos diretos de 3ª pessoa, são preferíveis a orações com objeto direto pronominal (cf. CYRINO, 1993, 1997). A literatura que relaciona ambos os fenômenos na história do PB mostra que houve uma mudança na seguinte direção: objetos nulos e sujeitos pronominais tiveram uma frequência crescente ao longo dos últimos séculos. Aqui defendemos uma relação ligando ambos os fenômenos no PB: ambos são sensíveis ao gênero semântico do referente quando se trata de realização de pronome de 3ª pessoa. Apoiamos tal hipótese trazendo dados de nossos trabalhos anteriores e de trabalhos de colegas. Também defendemos uma diferença crucial entre os dois fenômenos: embora objetos nulos estejam estabilizados na gramática do PB, apresentamos dados empíricos que mostram evidências de mudança no tempo aparente em sujeitos pronominais de 3ª pessoa. Coletamos e analisamos dados de um *corpus* oral contemporâneo (LínguaPOA) e aqui mostramos que (i) não existe mais a assimetria entre sujeitos de 1ª e 2ª pessoa (por um lado) e sujeitos de 3ª pessoa (por outro); (ii) sujeitos nulos de 1ª e 2ª pessoa são fenômenos estabilizados; e (iii) sujeitos nulos de 3ª pessoa ainda estão em processo de mudança.

Palavras-chave: mudança; sujeito nulo; sujeito pronominal; objeto direto anafórico; português brasileiro.

1 Introduction

Pronominal and null subjects and anaphoric direct objects are amongst the main grammatical phenomena that differentiate Brazilian and European varieties of Portuguese (cf. CYRINO; MATOS, 2016; DUARTE; FIGUEIREDO SILVA, 2016). In Brazilian Portuguese (BP), clauses with overt referential pronominal subjects are preferred over clauses with null subjects – which are restricted to some contexts (cf. AYRES, 2021; DUARTE; REIS, 2018), thus (1) is 'preferred' (or more frequent) in comparison to (2).

(1) A Maria acabou   de sair.   Ela estava bem   feliz.

the Mary finished of leave. She was   pretty happy

'Mary just left. She was pretty happy'

(2) A Maria acabou   de sair.   Ø Estava bem   feliz.

the Mary finished of leave.   Was   pretty happy

'Mary just left. She was pretty happy'

On the other hand, clauses with null anaphoric direct objects, especially 3rd person direct objects (3), are preferred over clauses with pronominal direct object (4) (cf. CYRINO, 1993, 1997; SCHWENTER, 2006; SCHWENTER; SILVA, 2003, *inter alia*).

(3) Comprei   um livro, mas ainda não tive   tempo de ler Ø.

Bought.1p.sg a book, but still   no had.1p.sg time of read

'I bought a book, but I still didn't have time to read it.'

(4) ?Comprei   um livro, mas ainda não tive   tempo de ler ele.

Bought.1p.sg a   book, but still   no had.1p.sg time of read he

'I bought a book, but I still didn't have time to read it.'

Literature has related (albeit timidly) both phenomena in the history of BP. An example of such a study is Tarallo's (1983, 1986) pioneering work, which indicates, through corpus analysis, a change in the expression of pronominal subjects and direct objects in BP. Tarallo highlights the relationship between the two phenomena, presenting diachronic data that show the increase of overt pronominal subject and the simultaneously decrease of pronominal direct objects in BP.

Graph 1 - Overt pronominal subjects and direct objects in BP across time.

Source: adapted from Tarallo (2018, p. 40).

Even though both changes 'started together' in BP and they seem to be related, as Tarallo pointed out, only the null direct object is considered a grammatical process whose change is stabilized (cf. CYRINO, 1993, 1997); the change in the parameter of subject marking in BP (from a +pro-drop to a -pro-drop language or, alternatively, to a partial null subject language) is still topic of lively debate among linguists who investigate the history of BP (cf. DUARTE; MARINS, 2021; DUARTE; REIS, 2018; GRAVINA, 2014a, 2014b; HOLMBERG; NAYUDU; SHEEHAN, 2009).

Here we have two main goals: the first goal is to argue for a relation connecting both phenomena in BP (null subjects and null objects); the second one is to argue for a crucial difference between them.

In section 2, we explore their connection. We believe that, since vernacular BP has lost 3rd person clitics (adopting the full pronouns *ele/ela/eles/elas*, 'he/she/they.masc./they.fem' along with the possibility of phonetically null objects), an agreement effect has become crucial to understand the distribution of pronouns and null objects. This agreement effect is related to the semantic gender of the referent, in the following sense: referents denoting masculine semantic gender (*o professor*, 'the male-teacher', for example) will favor pronominal direct objects with the masculine pronoun (*ele*, 'he'), referents denoting female semantic gender (*a professora*, 'the female-teacher', for example) will favor pronominal direct objects with the feminine pronoun (*ela*, 'she') and, finally, referents denoting neutral semantic gender (*a mesa*, 'the table', or *o livro*, 'the

book') will favor null objects, regardless of their grammatical gender (in BP, nouns can be masculine or feminine, *a mesa* being a feminine noun, *o livro*, a masculine noun).

This idea is not new and is not ours. It dates back to Creus and Menuzzi (2004) and has been explored by recent work (we provide references in section 2). Our idea here is to expand the 'semantic gender hypothesis' to investigate null and pronominal subjects. We believe this agreement effect is influencing both phenomena, pronominal direct objects and pronominal subjects, in the sense that a referent denoting masculine or feminine semantic gender will favor the use of a pronoun (in the way previously outlined), whereas a referent denoting neutral semantic gender will favor null subjects and null objects. In the next section, we will not present new data to support this idea; instead, we bring data from previous works (see references in the next section). This agreement effect is, thus, something connecting these two phenomena.

In section 3, we take a different direction. We explore a crucial difference between null subjects and null objects. As we have pointed out, Tarallo's early investigation showed that both pronominal direct objects and null subjects started to decline in the same period of time in BP history (mid-19th century). Nonetheless, we argue here that, in spite of the stabilization of null and pronominal direct objects, the null subject phenomenon is still a process in change in the grammar of the language. This idea is not new and is not ours either. Duarte (in many works), for example, advocates for the idea that null and pronominal subjects in BP are in constant change since the mid-19th century. Null subjects are residual and are diminishing both in frequency and in the variety of pragmatic and syntactic contexts where they are can appear. In other words, this grammatical phenomenon has not yet stabilized in the grammar. Here we bring new unpublished data to support the hypothesis of change. We have investigated a contemporary oral corpus, the LínguaPOA corpus, and found interesting results when we stratified the informants by age groups: we found that the older group (60+) uses considerably more 3rd person null subjects (46% of null subjects, 54% of pronominal subjects) than the younger group (20-39, where we found 11% off null subjects, 89% of pronominal subjects). That 'change-in-apparent-time' analysis supports Duarte's findings about the continuing change involving null and pronominal subjects in BP.

2 Null and overt pronominal subject and anaphoric direct object in BP, a correlation: the semantic gender hypothesis

Both null and overt pronominal subjects and anaphoric direct objects are strongly related (or so we argue) to 'natural' or 'semantic' gender agreement: BP has a masculine 3rd person pronoun that refers to male beings and referents ('ele', he), a feminine 3rd person pronoun that refers to female beings ('ela', she) and an empty category that refers to neutral beings and objects (Ø, it), a parallel, so to speak, to the English pronominal system, for example.

All nouns in Portuguese are either masculine or feminine in terms of *grammatical gender* (cf. CAMARA Jr., 2019, 2021), despite referring to inanimate (5) or animate referents (6). They agree in grammatical gender with other elements in the noun phrase, such as determiners and adjectives that modify them:

(5) o   livro caro   / a   casa   cara

the.masc book expensive.masc / the.fem house expensive.fem

'the expensive book / the expensive house'

(6) o   menino esperto   / a   menina esperta

the.masc boy   clever.masc / the.fem girl   clever.fem

'the clever boy / the clever girl'

Nouns that refer to inanimate entities, such as *livro* ('book'), *casa* ('house'), etc., or animate entities, such as *cônjuge* ('spouse'), *vítima* ('victim'), which have grammatical gender, but not semantic gender will be 'neutral', i.e. -sem.g. On the other hand, nouns with semantic gender (+sem.g) will either refer to feminine entities, such as *menina* ('girl'), *gata* ('female cat'), or masculine entities, *menino* ('boy'), *gato* ('cat'), and denote animate entities only.

We argue here (following COELHO *et al.*, 2017; CREUS; MENUZZI, 2004; OTHERO; SPINELLI, 2019a, 2019b) that the pronoun system in BP (at least when it comes to pronoun use as subjects or anaphoric direct objects) is sensitive to the 'semantic' gender of the referents. As we said in the previous section, we do not bring new evidence here nor any new theoretical point of view to support this hypothesis, but we present data from our previous investigations and from the works of colleagues.

Table 1 - Distributional tendency of overt and empty pronominal forms in contemporary Brazilian Portuguese.

Semantic gender of referent	Subject and object pronoun
feminine	ela
masculine	ele
neutral	Ø

Source: the authors.

The situation is much clearer when these pronominal forms are anaphoric direct objects. As Cyrino (1993, 1997) and Tarallo (1986) show, the use of 3rd person clitic pronouns *-o/-a/-os/-as* ('him/her/them.masc/ them.fem') as anaphoric direct objects in Brazilian Portuguese started to decline in mid-19th century (these clitic pronouns are accusative pronouns, still used in monitored speech by people with formal education in Brazil, as well as in high standard written language). With the decline of these accusative pronominal clitics in vernacular BP, the grammar explored two other strategies for 3rd person anaphoric direct objects: (i) the use of the nominative 'full' pronouns *ele/ela/eles/elas* ('he/she/they.masc/they. fem') and (ii) the null object Ø, as we can see in the following examples (see also discussion in Nunes, 2018, and Othero; Cardozo, 2017).

(7) Encontrei   -o   ontem.
    Found.1p.sg. him/it yesterday.
    'I found him/it yesterday'.

(8) Encontrei   ele   ontem.
    Found.1p.sg. he/it yesterday.
    'I found him/it yesterday'.

(9) Encontrei Ø ontem.
    Found.1p.sg. yesterday.
    'I found it yesterday'.

The original hypothesis that the semantic gender is guiding the distribution of nominative pronouns (8) and null objects (9) comes from

Creus and Menuzzi (2004). As the authors argue, this distribution is guided by an agreement principle (CREUS; MENUZZI, 2004, p. 161)[1]:

> The hypothesis that associates BP full pronouns to the presence of semantic gender, and null objects to the absence of semantic gender is more natural than the analogous hypothesis based on the distinction of animacy: after all, the basic difference between *ele/ela* ('he'/'she') and null objects is that the former carry gender specifications, while the latter are just unspecified for gender (…). In other words, the choice between null objects and full pronouns would basically result from a process of agreement between the antecedent and the anaphoric form: antecedents with semantic gender favor the use of pronouns because these are the anaphoric forms specified for gender; and antecedents without semantic gender favor the use of null objects precisely because null objects do not have a specification for semantic gender.

Thus, null objects point to referents with no semantic gender, whereas pronominal anaphoric objects preferably point to referents with semantic gender. The pronoun is a way of morphologically (and, thus, phonetically) exponentiating this semantic gender feature. Recent literature has supported this hypothesis. For example, Othero *et al.* (2016) conducted an experiment following this hypothesis and also the remark by Schwenter, to whom BP "displays a 'split' system of marking anaphoric DOs [direct objects]. All animate (mainly human) and specific anaphoric DOs are preferentially encoded overtly, while all others are preferentially encoded as null objects" (Schwenter, 2006, p. 26). This

---

[1]   Original text: "a hipótese que associa os pronomes plenos do PB à presença de gênero semântico, e objetos nulos à ausência de gênero semântico é mais natural que a hipótese análoga baseada na distinção de animacidade: afinal, a diferença básica entre as formas ele/ela e os objetos nulos é que as primeiras portam especificações de gênero, enquanto que os últimos são justamente não-especificados para gênero (bem como para número, mas nisso os ONs não diferem significativamente dos PrPls, já que os últimos podem ou não portar a flexão de número). Ou seja, a escolha entre ONs e PrPls resultaria, basicamente, de um processo de concordância entre antecedente e forma anafórica: antecedentes com gênero semântico favorecem o uso de PrPls porque estas são as formas anafóricas especificadas para gênero; e antecedentes sem gênero semântico favorecem o uso de ONs precisamente porque ONs não possuem especificação para gênero semântico."

'split system' can be explained by the semantic gender of the referents, as Othero *et al.* (2016) show. According to the authors (2016, p. 10)[2],

> What our data indicate quite clearly *does not concern the conditioning of the null object itself*, but the conditioning of the use of the full pronoun in the anaphoric resumption of antecedents with semantic gender, [+sem.g]. That is, analyzing the results of our tests, we can see that, if an antecedent has the feature [+sem.g], there is a strong tendency for it to be referred to by a pronoun – and not by a null object (…). On the other hand, if an antecedent does not have semantic gender, its resumption is sometimes made with a pronoun, sometimes with an empty category. This indicates, among other things, that there is a tendency towards complementary distribution, but it is not categorical.

In a later study, Othero and Spinelli (2017) analyzed a speech corpus (VARSUL[3], from the Rio Grande do Sul sample) from the 1990s and found more empirical data to support this hypothesis. Here we summarize some of their findings:

Table 2 - Null and pronominal objects in a spoken language corpus from the 1990s (3rd person only).

	+sem.g referents	-sem.g referents
Null objects	10/218 (4,5%)	**208/218 (95,4%)**
Pronominals	**41/61 (67,3%)**	20/61 (32,7%)

Source: adapted from Othero and Spinelli (2017, p. 188).

_____

[2] Original text: "O que nossos dados apontam de maneira bastante clara *não diz respeito ao condicionamento do objeto nulo* em si, mas ao condicionamento do uso do pronome pleno na retomada anafórica de antecedentes *com gênero semântico*, [+gs]. Ou seja, analisando os resultados de nossos testes, podemos perceber que, se um antecedente tem o traço [+gs], há uma forte tendência para que ele seja retomado por um pronome – e não por um objeto nulo (...). Por outro lado, se um antecedente não tiver o gênero semântico marcado, sua retomada ora é feita com um pronome, ora com uma categoria vazia. Isso indica, entre outras coisas, que há uma tendência à distribuição complementar, mas ela não é categórica."

[3] VARSUL is a corpus project that collects speech data, in the form of sociolinguistic interviews, from the southern region of Brazil, cf. Collischonn and Monaretto (2012) and Bisol and Monaretto (2016).

As the authors show, null objects tend to refer to antecedents with no semantic gender (95,4% of the cases), whereas full pronouns tend to refer to antecedents with apparent semantic gender (41 out of the 61 occurrences).

Coelho *et al.* (2017) also analyzed the VARSUL corpus, but from a different region (the sample from Santa Catarina), from the 1990s and 2010s. They reached a similar conclusion (2017, p. 2615)[4]:

> After reanalyzing all 322 occurrences of anaphoric direct objects with full pronouns (20 occurrences) and null forms (302 occurrences) in our sample, taking into account the three features of the antecedent NP (animacy, specificity and semantic gender), we conclude that the feature of semantic gender acts as a conditioning factor for the phenomenon: antecedents that have semantic gender favor the anaphoric resumption with the full pronoun; antecedents that do not have semantic gender favor the resumption with null object. (…)
>
> The semantic gender feature, in fact, seems to act not only on the phenomenon of anaphoric direct object of 3rd person, but on the phenomenon of anaphoric direct object in general in BP (involving 1st, 2nd and 3rd grammatical persons, both in pronominal as null objects), as suggested by the works of Schwenter (2006) and Othero et al. (2016) (…).

---

4   Original text: "Após reanalisar todas as 322 ocorrências de retomadas anafóricas com pronomes plenos (20 ocorrências) e objetos nulos (302 ocorrências) em nossa amostra, levando em consideração os três traços do SN antecedente (animacidade, especificidade e gênero semântico), concluímos que o traço de gênero semântico atua como fator condicionador para o fenômeno: os SNs antecedentes que têm gênero semântico identificado favorecem a retomada anafórica com o pronome pleno; os SNs antecedentes que não têm gênero semântico identificado favorecem a retomada com o objeto nulo (...)

O traço de gênero semântico, na verdade, parece atuar não somente sobre o fenômeno da retomada de objeto direto anafórico de 3ª pessoa, mas sobre o fenômeno de retomada anafórica de objeto direto em geral em PB (envolvendo pronomes e ONs, de 1ª, 2ª e 3ª pessoas gramaticais), como sugerem os trabalhos de Schwenter (2006) e Othero et al. (2016) (...)."

The authors go further (2017, p. 2615)[5]:

> If this is indeed the case, we can expect that this feature (semantic gender) is actually behind the distribution of pronouns and empty elements not only in the anaphoric direct object, but also in the subject position. This would allow us to reach an interesting generalization: that referents marked with semantic gender would favor the use of full pronouns in BP, whether as direct objects or as subjects.

Following this suggestion, Othero and Spinelli (2019b) analyzed null and pronominal *subjects* in a spoken language corpus from the 1990s (VARSUL, Porto Alegre sample). They found the same tendency: overt pronominal subjects tended to refer to +sem.g referents, whereas null subjects tended to refer to -sem.g referents. Some of their results can be seen in Table 3:

Table 3 - Null and overt pronominal subjects according to the semantic gender feature of their referent in spoken language corpus from the 1990s.

	+sem.g referents	-sem.g referents
Pronominal subjects	**250/336 (74,4%)**	86/336 (25,6%)
Null subjects	32/106 (30,2%)	**74/106 (69,8%)**

Source: Adapted from Othero and Spinelli (2019b, p. 19).

In the same study, the authors investigated a more recent spoken language corpus from the same location as VARSUL, the LínguaPOA corpus, containing sociolinguistic interviews from 2015 to 2018[6]. They found very similar results, strengthening the hypothesis mentioned by Coelho *et al.* (2017, p. 2615), "that this feature (semantic gender) is actually behind the distribution of pronouns and empty elements not

---

[5] Original text: "Se esse for realmente o caso, podemos esperar que esse traço (gênero semântico) esteja por trás, na verdade, da distribuição entre pronomes *versus* elementos vazios não apenas na retomada anafórica de objeto direto, mas também na função de sujeito. Isso nos permitiria chegar a uma generalização interessante: a de que os referentes marcados com gênero semântico favoreceriam o uso de pronomes retos em PB, seja na função de objeto direto, seja na função de sujeito."

[6] LinguaPOA is an oral corpus composed of sociolinguistic interviews with informants from the city of Porto Alegre, cf. Battisti (2019).

only in the anaphoric direct object, but also in the subject position". In the table below, we report Othero and Spinelli's (2019b) results:

Table 4 - Null and overt pronominal subjects according to the semantic gender feature of their referent in a corpus from the 2010's.

	+sem.g referents	-sem.g referents
Pronominal subjects	**236/317 (74,4%)**	81/317 (25,6%)
Null subjects	8/86 (9,3%)	**78/86 (90,7%)**

Source: Adapted from Othero and Spinelli (2019b, p. 22).

However, there are more factors playing a role in the distribution of null and overt pronominal subjects in BP than just the semantic gender feature of the referent. Other factors related to the phenomenon are topic-chain continuity (cf. LAZZARI, 2021; PAREDES SILVA, 2003), linear order or the elements in the clause (cf. KATO, 2020; KATO; DUARTE, 2018) and verbal inflectional morphology (cf. SOARES, 2017; SOARES; MILLER; HEMFORTH, 2019). Our point here is that the semantic gender hypothesis can be applied to the investigation of pronominal subjects and can (help) explain the distribution of 3rd person pronominal direct objects in BP. It is certainly not the only factor at play, but it is certainly one of them (cf. Ayres, 2021, Othero; Spinelli, 2019a, 2019b, Soares; Miller; Hemforth, 2020, for a broad analysis and argumentation). For our purposes here, this is important because unlike topic-chain continuity, linear order and verbal inflection morphology (that are related to the distribution of null and pronominal subjects – but *not* to pronominal direct objects), the semantic gender feature of the referent is a factor playing a role both in the distribution of null and pronominal subjects *and* null and pronominal direct objects in BP. In other words, the semantic gender feature can be pointed out as the factor that connects both phenomena in BP, since we are facing an agreement effect linking pronouns and referents.

We did not present any new data in this section. What we did was bring data from the literature (ours and from colleagues) attesting two points of connection between null subjects and null objects in BP, namely, (i) their frequency started changing in the same period of BP's history, around mid-19th century, and (ii) their sensitivity to semantic gender information of referents, in the sense that referents denoting neutral

semantic gender favor null forms, whereas masculine and feminine referents (in terms of semantic gender – not grammatical gender) favor pronominal forms.

In the next section, we will explore a crucial difference between the two phenomena. We will present new data to support the idea that the null subject is still changing in contemporary BP. We present a quantitative analysis[7] of the data we found investigating 5,846 occurrences of null and pronominal subjects in the interviews from the sociolinguistic corpus LínguaPOA. We show that there are elements that allow us to pursue the idea of a change in apparent time when it comes to the null subject phenomenon.

3 Change

Although the two phenomena are closely related, as we explored in the previous section, and although null object is a stabilized grammatical phenomenon in the grammar of BP since the late 20th century, we argue here that the null subject phenomenon is still a grammatical process in change. As we have already mentioned, this idea is not new and is not ours. What we do here is bring new data to support this claim. We have analyzed a contemporary oral corpus of vernacular BP, the LínguaPOA corpus (cf. BATTISTI, 2019). The corpus is constituted by a series of transcribed sociolinguistic interviews from 2015 to 2018 with informants from the city of Porto Alegre, a capital city in Southern Brazil.

We have analyzed 5,846 occurrences of null and pronominal subjects and found the following distribution:

Table 5 – Null vs. overt pronominal subjects found in the corpus.

| Pronominal subjects	4,220 (72%)
Null subjects	1,626 (28%)
Total	5,846 (100%)

Source: the authors.

---

[7]   We present a qualitative analysis of the data in Ayres and Othero (forthcoming) and Othero, Lazzari and La Porta (forthcoming).

The data we found, at a first glance, lead us to think that the null subject in BP has stabilized in the late 20th century (the same period of time we find the stabilization of null objects), since we found roughly the same figures that have been long reported in the literature, i.e., around 70% of the anaphoric subjects being filled with an overt pronoun and 30% of null subjects (cf. BERLINCK; DUARTE, 2015, DUARTE, 1995, MONTEIRO, 1994).

Nonetheless, we have made three interesting empirical discoveries in our investigation. The first two are related to 1st and 2nd person subjects versus 3rd person subjects. The data from the 1990s present an asymmetry for 3rd person subjects, when compared to 1st and 2nd persons: Duarte (1993, 1995), analyzing theater plays, found that 1st and 2nd person subjects were majorly overt, whereas 3rd person subjects were split: roughly 45% were pronominal subjects and roughly 55% were null subjects (cf. DUARTE, 1993, p. 117):
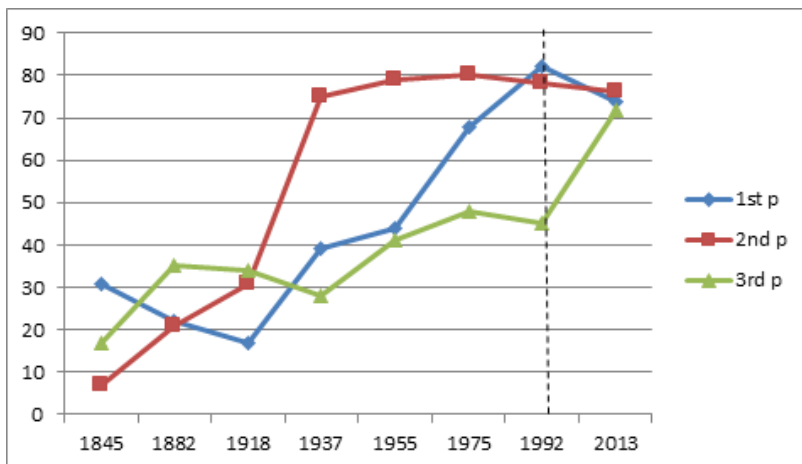
Graph 2 - Null vs. overt pronominal subjects in theatre plays.
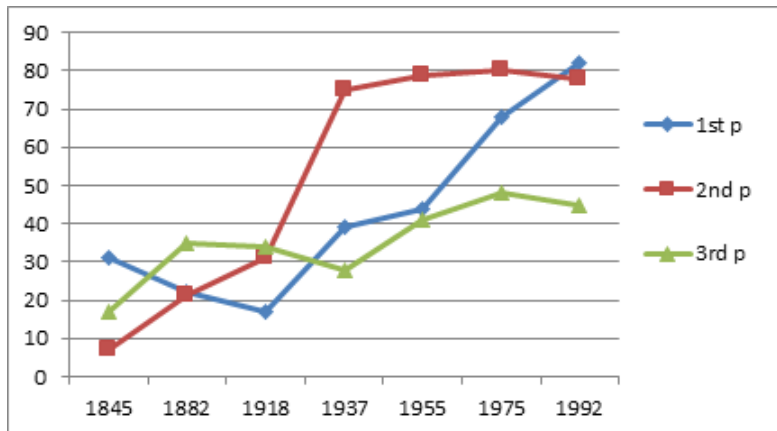
Source: adapted from Duarte (1993, p. 117).

In a recent work, Othero and Spinelli (2019a) investigated theatre plays from 2010s (thirty years after the last play analyzed by Duarte) and found out that the asymmetry reported by Duarte (1993) involving the 3rd person no longer exists.

Graph 3 - Null vs. overt pronominal subjects in theatre plays.

Source: adapted from Othero; Spinelli (2019a, p.16).

Here, in our investigation of 3rd person subjects in an oral corpus, we found similar results: overt pronominal subjects are preferred over null subjects, confirming what Othero and Spinelli (2019a) found when investigating their corpus of theater plays from 2010[8]. In our investigation of LínguaPOA, we found 715 (72%) occurrences of 3rd person pronominal subjects versus 281 (28%) of 3rd person null subjects, as we show in Graph 4.

Graph 4 - 3rd person null vs. overt pronominal subjects.

Source: the authors.

_____

[8]   Similar results have been reported by Berlink, Duarte and Oliveira (2015), who found 78% of 3rd person pronominal subjects in their oral corpus investigation.

That means we have found roughly the same distribution (~70/~30) between null and pronominal subjects across 1st, 2nd and 3rd persons, and the asymmetry between 3rd person pronominal/null subjects and 1st and 2nd person subjects no longer exists. The lack of this long-reported asymmetry was the first interesting finding from our corpus investigation, even though it may also lead us to think we are now facing a stable process in the grammar of BP, since all persons (1st, 2nd and 3rd), after all, present nearly the same frequency for overt pronominal subjects – as we show in Table 7:

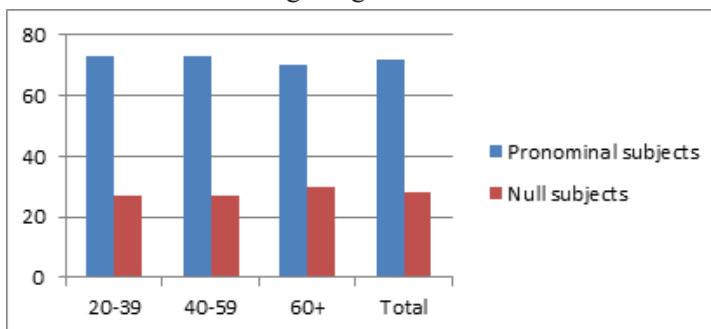Table 7 – Null vs. overt pronominal subjects found in the corpus.

	1st and 2nd person sg.	3rd person sg.
Pronominal subjects	3,505 (72%)	715 (72%)
Null subjects	1,345 (28%)	281 (28%)
Total	4,850 (100%)	996 (100%)

Source: the authors.

The second finding is another piece of empirical data that could support the idea of nulls subjects being a stable phenomenon in the grammar. It appeared when we decided to stratify the informants according to age group. LínguaPOA staff collected personal information from the informants, and we used the three age groups established by the LínguaPOA project: 20-39 years old, 40-59 years old and 60+ years old. When we analyzed 1st and 2nd person subjects, we found nearly the same distribution, as we show in Graph 5 (in percentages) and Table 8:

Graph 5 – 1ˢᵗ and 2ⁿᵈ person singular, null vs. overt pronominal subjects, according to age stratification.
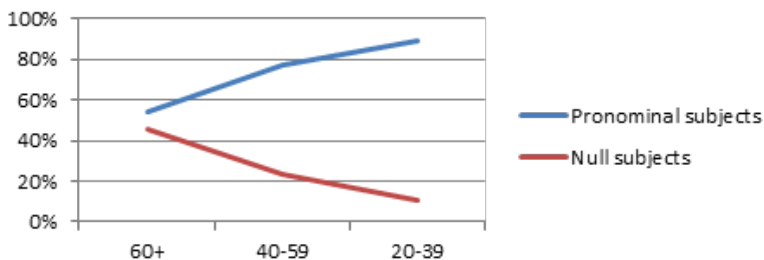

Source: the authors.

Table 8 – 1ˢᵗ and 2ⁿᵈ person singular, null vs. overt pronominal subjects, according to age stratification.

Age group	20-39	40-59	60+
Pronominal subjects	1,489 (73%)	1,189 (73%)	827 (70%)
Null subjects	537 (27%)	450 (27%)	358 (28%)
Total	2,026 (100%)	1,639 (100%)	1,185 (100%)

Source: the authors.

That means, to our understanding, we have evidence – both from diachronic analyses (such as Duarte, 1993, 1995, 2018) and from our own synchronic analysis – that the null subject is stable *when it comes to 1ˢᵗ and 2ⁿᵈ person singular in BP*. On the other hand, when the 3ʳᵈ person singular is concerned, things are different – here is our third, and crucial, finding. When we analyzed 3ʳᵈ person null and pronominal subjects according to age group, we found an interesting distribution that, to our understanding, is evidence of change in apparent time. We present our findings in Graph 6. The horizontal axis corresponds to age stratification (in years); the vertical axis to the application of the rule (in percentage), with the blue line marking the frequency of pronominal subjects and the red line marking the frequency of null subjects.

Graph 6 – 3rd person null and overt pronominal subjects according to age stratification.
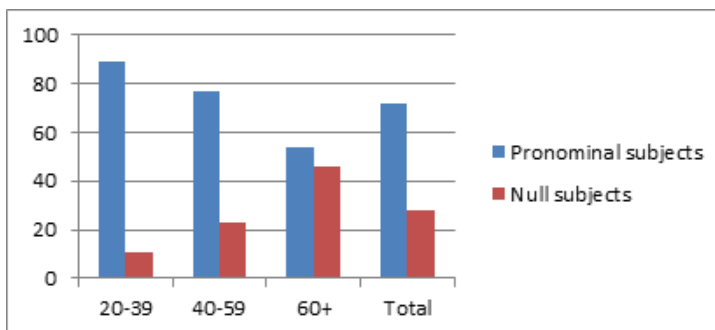

Source: the authors.

The older group presents a close distribution in the occurrences of 3rd person null and pronominal subjects, whereas the younger group has a clear preference for pronominal subjects. We show this data also in Table 9 and Graph 7:

Table 9 – 3rd person singular, null vs. overt pronominal subjects, according to age stratification.

Age group	20-39	40-59	60+
Pronominal subjects	187 (89%)	362 (78%)	175 (54%)
Null subjects	22 (11%)	101 (22%)	149 (46%)
Total	209 (100%)	463 (100%)	324 (100%)

Source: the authors.

Graph 7 – 3rd person null vs. overt pronominal subjects according to age stratification.

Source: the authors.

If we consider only the total distribution of null and pronominal subjects in the corpus (the last pair of columns in Graph 7), disregarding the age factor, we are misled to think the numbers are virtually the same for 1st, 2nd and 3rd person distribution between pronominal and null subjects (~70/~30, cf. Graph 5). However, analyzing the data more closely and relating the phenomena to the age factor, it seems clear that we are facing a process of change, supporting the argumentation in Duarte and Marins (2021), for example, who have reached a similar conclusion analyzing a very different corpus. We need, however, to keep further investigation in order to find more occurrences of null and pronominal subjects. We should also analyze different kinds of corpora, from different regions of Brazil, with spoken and written varieties.

4 Final remarks

We had two goals here: the first was to review some of the literature relating the phenomena of null and overt pronominal subjects and direct objects in Brazilian Portuguese. We proposed an approximation of these two phenomena via the 'semantic gender hypothesis', basing our argumentation on our own previous works and on colleagues' investigations.

The second goal was to present new empirical data from our investigation of a contemporary transcribed BP oral corpus, the LínguaPOA corpus. What we found and presented in section 3 supports

the ideas that (i) the asymmetry between 1st and 2nd person subjects on the one hand and 3rd person subjects on the other no longer exists; (ii) 1st and 2nd person null subjects are a stabilized grammatical phenomena, since the frequency of null subjects has not changed since the late 20th century and it does not present any variation among different age groups; and (iii) 3rd person null subjects are still changing, since the data show change in apparent time.

Acknowledgments

Authorship statement

Gabriel de Ávila Othero analyzed the data and wrote the majority of the paper. Melissa Lazzari analyzed the data and revised the paper.

References

AYRES, M. R. *Contextos licenciadores de sujeitos nulos em português brasileiro*. 2021. 123f. Tese (Doutorado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, 2021.

AYRES, M. R.; OTHERO, G. A. *Contexts for Null Subjects in Contemporary Brazilian Portuguese*. Forthcoming.

BATTISTI, E. *O acervo de entrevistas sociolinguísticas LínguaPOA:* constituição, possibilidades e desafios. Rio de Janeiro: e-book Open Access, 2019.

BERLINK, R. A.; DUARTE, M. E. L.; OLIVEIRA, M. Predicação. *In*: KATO, M. A.; NASCIMENTO, M. (orgs.). *A Construção da Sentença:*

Gramática do Português Culto Falado no Brasil. São Paulo: Contexto, 2015. p. 81-150.

BISOL, L.; MONARETTO, V. N. O. Prefácio: VARSUL e suas origens, uma história sumariada. *ReVEL*, Novo Hamburgo, edição especial, n. 13, p. 6-11, 2016.

CAMARA JR. *Problemas de linguística descritiva*. Edição revista e comentada. Petrópolis: Vozes, 2021.

CAMARA JR., J. M. *Estrutura da língua portuguesa:* edição crítica. Edição, estabelecimento de texto, introdução e notas de Emílio Gozze Pagotto, Maria Cristina Figueiredo Silva, Manoel Mourivaldo Santiago-Almeida. Petrópolis: Vozes, 2019.

COELHO, I. L.; OTHERO, G. A.; VIEIRA-PINTO, C. A. Reanálise de variáveis semânticas no condicionamento do objeto nulo e do pronome pleno na fala de Florianópolis. *Fórum Linguístico*, Florianópolis, v. 14, n. 4, p. 2606-2617., 2017. DOI: https://doi.org/10.5007/1984-8412.2017v14n4p2606

COLLISCHONN, G.; MONARETTO, V. Banco de dados VARSUL: a relevância de suas características e a abrangência de seus resultados. *ALFA: Revista de Linguística*, v. 56, n. 3, p. 835-853, 2012.

CREUS, S; MENUZZI, S. O papel do gênero na alternância entre objeto nulo e pronome pleno em português brasileiro. *Revista da ABRALIN*, v. 3, n. 1-2, p. 149-176, 2004.

CYRINO, S. *et al. História do português brasileiro:* mudança sintática do português brasileiro: perspectiva gerativista. São Paulo: Contexto, 2018.

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: objeto nulo e clíticos. *In*: ROBERTS, I.; KATO, M. A. (orgs.) *Português brasileiro:* uma viagem diacrônica. Campinas: Ed. da Unicamp, 1993. p. 163-185.

CYRINO, S. M. L.; MATOS, G. Null objects and VP ellipsis in European and Brazilian Portuguese. *In*: WETZELS, L.; COSTA, J.; MENUZZI, S. (eds) *The handbook of Portuguese linguistics*. Oxford: Blackwell, 2016. p. 294-317.

DUARTE, I.; FIGUEIREDO SILVA, M. C. The null subject parameter and the structure of the sentence in European and Brazilian Portuguese. *In*: WETZELS, L.; COSTA, J.; MENUZZI, S. M. (eds.). *The handbook of Portuguese Linguistics*. West Sussex: Wiley Blackwell, 2016. p. 234-253.

DUARTE, M. E. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. *In*: ROBERTS, I.; KATO, M. A. (orgs.). *Português brasileiro:* uma viagem diacrônica. Campinas: Ed. da Unicamp, 1993. p. 107-125.

DUARTE, M. E. L. *A perda do princípio "evite pronome" no português brasileiro*. 1995. 151 f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Unicamp, 1995.

DUARTE, M. E. L. O sujeito nulo no português brasileiro. *In*: CYRINO, S.; MORAIS, M. A. T. *História do português brasileiro.* Mudança sintática do português brasileiro: perspectiva gerativista. São Paulo: Contexto, 2018. p. 26-71.

DUARTE, M. E. L.; MARINS, J. Brazilian Portuguese: a 'partial' null subject language. *Cadernos de Estudos Linguísticos*, Campinas, v. 63, p. 1-21, 2021. DOI: https://doi.org/10.20396/cel.v63i00.8661660

DUARTE, M. E. L.; REIS, E. P. R. Revisitando o sujeito pronominal vinte anos depois. *ReVEL*, Novo Hamburgo, v. 16, n. 30, p. 173-197, 2018.

GRAVINA, A. P. Diacronia e sujeito nulo no português brasileiro: um estudo comparativo. *Filologia e Linguística Portuguesa*, São Paulo, v. 16, p. 199-231, 2014a. DOI: https://doi.org/10.11606/issn.2176-9419.v16ispep199-231

GRAVINA, A. P. *Sujeito nulo e ordem VS no português brasileiro: um estudo diacrônico-comparativo baseado em corpus*. 2014. 251 f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Unicamp, 2014b.

HOLMBERG, A.; NAYUDU, A.; SHEEHAN, M. Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica*, Lund, v. 63, n. 1, p. 59-97, 2009. DOI: https://doi.org/10.1111/j.1467-9582.2008.01154.x

KATO, M. A.; DUARTE, M. E. L. Pre-verbal position in BP: a reinterpretation of "avoid pronoun principle". *Diadorim*, Rio de Janeiro,

v. 20 – Especial, p. 610-626, 2018. DOI: https://doi.org/10.35520/ diadorim.2018.v20n0a23293

KATO, M. *Determinantes prosódicos em mudança sintática*. Abralin ao vivo, 2020. <http:// https://youtu.be/t3BLRPloZJI >. Acesso em: 12 nov. 2021.

LAZZARI, M. G. Acessibilidade: o que isso tem a ver com o sujeito pronominal expresso e o sujeito nulo em português brasileiro? *Revista Linguística Rio*, Rio de Janeiro, v. 7, p. 42-58, 2021.

MONTEIRO, J. L. *Pronomes pessoais:* subsídios para uma gramática do português do Brasil. Fortaleza: Edições UFC, 1994.

NUNES, J. Direção de cliticização, objeto nulo e pronome tônico na posição de objeto em português brasileiro. *In*: ROBERTS, I.; KATO, M. (orgs.). *Português brasileiro:* uma viagem diacrônica. São Paulo: Contexto, 2018. p. 161-174.

OTHERO, G. A.; AYRES, M. R.; SCHWANKE, C.; SPINELLI, A. C. A relevância do traço gênero semântico na realização do objeto nulo em português brasileiro. *Working Papers em Linguística,* Florianópolis, v. 17, n. 1, p. 64-86, 2016. DOI: http://dx.doi.org/10.5007/1984-8420.2016v17n1p64

OTHERO, G. A.; CARDOZO, R. W. A ordem pronominal em português brasileiro: da ênclise à próclise, do clítico ao tônico (or There and Back Again, a Word Order's Holiday). *Fórum Linguístico*, Florianópolis, v. 14, n. 1, p. 1717-1734, 2017. DOI: https://doi.org/10.5007/1984-8412.2017v14n1p1717

OTHERO, G. A.; LAZZARI, M. G.; LA PORTA, B. *Contextos de resistência do sujeito nulo na análise do corpus LínguaPoa*. Forthcoming.

OTHERO, G. A.; SPINELLI, A. C. Sujeito pronominal expresso e nulo no começo do séc. XXI (e sua relação com o objeto nulo em PB). *Domínios de Lingu@gem*, Uberlândia, v. 13, n. 1, p. 7-33, 2019a. DOI: https://doi. org/10.14393/DL37-v13n1a2019-1

OTHERO, G. A.; SPINELLI, A. C. Um tratamento unificado da omissão e da expressão de sujeitos e objetos diretos pronominais de 3ª pessoa em português brasileiro. *Cadernos de Estudos Linguísticos*, Campinas, v.61, n. 1, p. 1-30, 2019b. DOI: https://doi.org/10.20396/cel.v61i1.8654211

PAREDES SILVA, V. L. Motivações funcionais no uso do sujeito pronominal: uma análise em tempo real. *In*: PAIVA, M. C.; DUARTE, M. E. L. (orgs.). *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa, 2003. p. 97-114.

PIVETTA, V. *Objeto direto anafórico no português brasileiro*: uma discussão sobre a importância dos traços semântico-pragmáticos – animacidade/especificidade vs. gênero semântico. 2015. 131f. Dissertação (Mestrado) – Faculdade de Letras, Universidade Federal do Rio Grande do Sul, 2015.

SCHWENTER, S. A. Null objects across South America. Selected proceedings of the 8th. *In*: HISPANIC LINGUISTICS SYMPOSIUM, 2006, Somerville: Cascadilla Press, 2006. p. 23-37.

SCHWENTER, S. A.; SILVA, G. Anaphoric direct objects in spoken Brazilian Portuguese: semantics and pragmatics. *Revista Internacional de Lingüística Iberoamericana,* Frankfurt\Madrid, v. 1, n. 2, p. 99-123, 2003.

SOARES, E. C. *Anaphors in discourse:* anaphoric subjects in Brazilian Portuguese. 2017. 482 f. Tese (Doutorado em Linguística) – Faculdade de Letras, University Sorbonne, 2017.

SOARES, E. C.; MILLER, P.; HEMFORTH, B. The effect of verbal agreement marking on the use of null and overt subjects. *Forum lingüístico*, Florianópolis, v.16, n.1, p. 3579-3600, 2019. DOI: https://doi.org/10.5007/1984-8412.2019v16n1p3579

SOARES, E. C.; MILLER, P.; HEMFORTH, B. The effect of semantic and discourse features on the use of null and overt subjects: a quantitative study of third person subjects in Brazilian Portuguese. *DELTA*, São Paulo, v. 36, p. 1-38, 2020. DOI: http://dx.doi.org/10.1590/1678-460X2020360107

TARALLO, F. *Relativization strategies in Brazilian Portuguese*. 1983. 270 f. Tese (Doutorado em Linguística) – Faculdade Letras, University of Pennsylvania, 1983.

TARALLO, F. Sobre a alegada origem crioula do português brasileiro: mudanças sintáticas aleatórias. *In*: ROBERTS, I.; KATO, M. A. (orgs.). *Português brasileiro:* uma viagem diacrônica. São Paulo: Contexto, 2018. p. 29-54.