

CONTEXTS FOR NULL SUBJECTS IN CONTEMPORARY BRAZILIAN PORTUGUESE*CONTEXTOS PARA SUJEITOS NULOS EM PORTUGUÊS BRASILEIRO CONTEMPORÂNEO*Mônica Rigo Ayres¹Gabriel de Ávila Othero²**ABSTRACT**

Null subjects have been extensively investigated in natural languages. Brazilian Portuguese (BP), in particular, raises a lot of interest in the subject, for it is apparently in a process of grammatical change, from a *pro-drop* language to a *partial pro-drop language* (cf. BIBERAUER *et al.*, 2010; HOLMBERG *et al.*, 2009; DUARTE; MARINS, 2021; SOARES *et al.*, 2019). In this article, we investigate null subjects in BP assuming that null referential subjects are only used in ‘marked’, specific contexts, whereas overt subjects are the ‘unmarked’, more frequent strategy. We have analyzed a contemporary corpus of spoken BP and found 1,252 occurrences of null subjects. We then present four factors that are traditionally known for playing a role in determining the optimal contexts for null subjects in BP (and several other languages). We argue here that all these four factors are both necessary and sufficient to explain all null subject occurrences in any given corpus of contemporary spoken BP. We analyzed all the 1,252 occurrences of null subjects in the corpus and could explain 99.8% of the data, leading us to a better understanding of the null subject phenomenon in BP, at least when it comes to identifying its licensing contexts.

KEYWORDS: Null subject. Corpus analysis. Brazilian Portuguese.

RESUMO

Sujeitos nulos têm sido largamente investigados em línguas naturais. O português brasileiro (PB), em particular, desperta muito interesse pelo assunto, pois aparentemente o PB está em processo de mudança, de uma língua *pro-drop* para uma língua *pro-drop* parcial (cf. BIBERAUER *et al.*, 2010; HOLMBERG *et al.*, 2009; DUARTE; MARINS, 2021; SOARES *et al.*, 2019). Neste artigo, investigamos sujeitos nulos em PB assumindo que sujeitos referenciais nulos são usados apenas em contextos “marcados”, específicos, enquanto sujeitos expressos são a estratégia “não marcada”, mais frequente. Analisamos um *corpus* contemporâneo de PB falado e encontramos 1.252 ocorrências de sujeitos nulos. Em seguida, apresentamos quatro fatores que são tradicionalmente conhecidos por desempenhar algum papel na determinação dos contextos ótimos para sujeitos nulos em PB (e várias outras línguas). Argumentamos aqui que todos esses quatro fatores são necessários e suficientes para explicar todas as ocorrências de sujeito nulo em qualquer *corpus* de PB falado contemporâneo. Analisamos todas as 1.252 ocorrências de sujeitos nulos no *corpus* e conseguimos explicar 99,8% dos dados, o que nos leva a uma melhor compreensão do fenômeno do sujeito nulo no PB, pelo menos que tange à identificação de seus contextos de licenciamento.

PALAVRAS-CHAVE: Sujeito nulo. Análise de *corpus*. Português Brasileiro.

¹ Universidade Federal do Ceará, monica.ayres@ufrgs.br, <https://orcid.org/0000-0003-0972-5633>.

² Universidade Federal do Rio Grande do Sul, gabriel.othero@ufrgs.br, <https://orcid.org/0000-0002-2060-6312>.

Introduction³

Null subjects have already been extensively investigated in natural languages. Brazilian Portuguese (BP), in particular, raises a lot of interest in the subject, for it is apparently changing from a pro-drop language to a partial pro-drop language (BIBERAUER *et al.*, 2010; HOLMBERG *et al.*, 2009, DUARTE; MARINS, 2021; SOARES *et al.*, 2019). Here, we will investigate null subjects in BP assuming that null referential subjects are only used in specific contexts, whereas overt subjects are the ‘unmarked’, more frequent strategy.

According to Widera; Kaiser (2019, p. 144), “Portuguese is considered to be a null subject language. This means that, in this language, the subject’s omission represents the normal case and the use of the subject pronoun is limited to restricted contexts” – that is true for European Portuguese; in Brazilian Portuguese, we find the exact opposite scenario (DUARTE, 1995; DUARTE, 2012; AYRES, 2021), i.e. in BP the subject’s omission represents the ‘marked’, less common case, and it is restricted to some specific contexts – which we intend to cover here.

Literature on BP null subjects (and on null subjects in general, across the languages) usually points to four factors that favor null subjects: (i) rich verbal AGR morphology (DUARTE, 1993); (ii) semantic features of the referent DP [human, referential] (CYRINO; DUARTE; KATO, 2000) or [semantic gender] (OTHERO; SPINELLI, 2019b); (iii) *V1 linear order (KATO, 2020); and (iv) optimal discourse connection (or topic chain) (PAREDES SILVA, 2003).

Here, in our investigation, we intend to analyze a BP spoken language corpus to verify if the null subjects we encounter can be explained by one or more of these four factors. In other words, we aim to check if null subjects are licensed in contexts where

- i. verbal AGR morphology is rich enough to explicit the subject;
- ii. the referent of the null subject is a [–human, –referential] DP (or a DP with the feature [–semantic gender]);
- iii. the verb is not the first element in the clause;
- iv. the null subject occurs in an optimal connection context, i.e. it is a prominent element in a topic chain.

In the next sections, we will explain in detail each of these contexts and features and their relations to null subject occurrences in BP. After that, we will present the data we have investigated and analyze the occurrences according to these four factors.

1. Null subjects in Brazilian Portuguese

The investigation of null subjects in natural languages, in the generative framework, dates back,

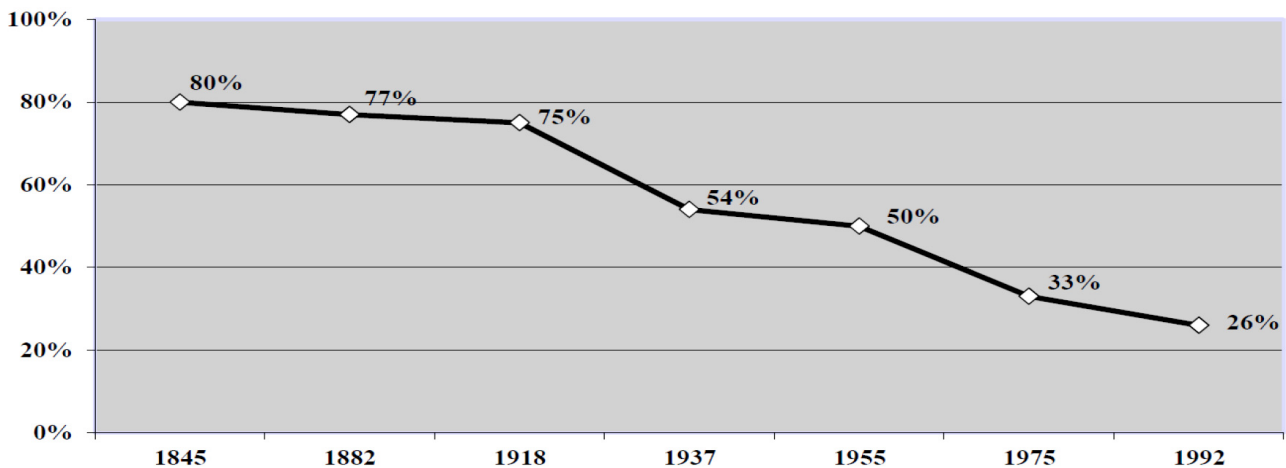
³ We would like to thank Aline Gravina, Camila Ulrich, Georg Kaiser, Karoline Gasque, Melissa Lazzari and Sergio Menuzzi for debating with us many of the ideas we present here. We would also like to thank Elisa Battisti, for sharing the LínguaPOA corpus with the academic community. Any inconsistencies in our text are of our own responsibility.

at least, to Perlmutter (1971), who pointed out the distinction between languages that allow null subjects and languages that do not. The topic received more attention after Chomsky (1981), when the pro-drop parameter was formalized. This formalization took place in the Principles and Parameters framework, which predicts that there are linguistic universals (principles), with parameters being the major responsible for differences among languages. For example, think about the Extended Projection Principle (EPP): this principle postulates that, in any natural language, every clause will have a subject. The value that the null subject parameter receives in each language will define whether or not this subject needs to be overt, i.e., morphologically and phonetically expressed. If the subject does not necessarily need to be expressed, we are facing a common phenomenon in natural languages: the null subject (1). If a language does not allow the omission of the subject, i.e., the phonetically expressed subject is mandatory, we say that language is a –pro-drop language (2).

- 1) (Io) Sono uno studente (Italian)
(Yo) Soy estudiante (Spanish)
- 2) *(I) am a student (English)
*(Je) suis un étudiant (French)

The picture, though, is not so clear as one might think, for there are languages that have some +pro-drop characteristics and –pro-drop as well. Brazilian Portuguese is such a language (cf. BIBERAUER *et al.*, 2010, DUARTE; FIGUEIREDO SILVA, 2016, HOLMBERG *et al.*, 2009). Tarallo (1983) was probably the first to investigate a possible change in the parameter setting in BP. Duarte (1993, 1995), observing this change, followed Tarallo’s suggestions and pointed out that BP goes through a transition phase from a +pro-drop to a –pro-drop language stating that “the increasingly less frequent occurrences of first person singular null subjects in oral language, still licensed by the verbal agreement morphology, should continue to manifest themselves indefinitely as residues of the pro-drop property” (DUARTE 1995, p. 56). Still in the mid-1990s, Monteiro (1994, p. 123), in a comprehensive investigation of pronouns in BP, questions if it is valid to consider BP as a null subject (+pro-drop) language, as it is the case for European Portuguese.

This question is well founded, as several studies have shown that the subject position/function is preferably overt in BP, as opposed to what used to happen in earlier periods of time – the change over time can be seen in the following graph:

Graph 1: Null subjects across time in Brazilian Portuguese

Source: DUARTE, 2018, p. 88

As we can see, null subjects went through a decrease: they reached 80% in the first period analyzed by Duarte (1845) and dropped to only 26% in the most recent period of Duarte (2018)' analyses. As Duarte (1995, p. 4) states: "although it cannot be said that we have lost the possibility of omitting the subject, there is a clear preference for using the full pronominal form".

When comparing BP to other +pro-drop Romance languages (European Portuguese, Spanish and Italian, cf. KATO, 1999), we see that the contexts in which null subjects occur are more restricted in BP than in these other languages. This is why BP has been classified by some as a partial pro-drop language (RODRIGUES, 2004; HOLMBERG, 2005; GRAVINA, 2014). According to Gravina (2014, p. 33), BP is a partial null subject language, since "it still licenses null subject, but with context restriction". Soares; Miller; Hemforth (2019, p. 3581) point to two different approaches:

Within the Generative tradition, there are two general approaches to so-called 'pro-drop' in BP, one diachronic, the other synchronic. These can be taken to be complementary. According to the diachronic approach, there is an ongoing change in the pro-drop parameter towards obligatory phonological realization of subject pronouns (TARALLO, 1983, among others); according to the synchronic approach, present-day BP is a partial pro-drop language (HOLMBERG *et al.*, 2009; BIBERAUER *et al.*, 2010, *inter alia*).

In this same text, a reservation is made regarding these two positions:

As Sérgio Menuzzi (p. c.) has pointed out, though these two positions do not exclude each other, they were not put forward as compatible in the literature and no specific framework within which they might form a single combined diachronic and synchronic hypothesis has been proposed. Specifically, researchers defending the 'partial pro-drop' analysis have mainly argued for this position against the idea that BP was becoming a non-null subject language of the English type. (SOARES; MILLER; HEMFORTH, 2019, p. 3581)

Partial null subject languages, though, do not behave exactly in the same way. Duarte (2020, p. 95) emphasizes that “the inclusion of BP among the partial null subject languages needs to be seen carefully. According to Biberauer (2010), this ‘partial’ label may contain very different systems, which require a rigorous analysis of the contexts that license null subjects.” Framing languages into categories is not a trivial task, for they do not behave in the same way in licensing null subjects.

In any case, we will not discuss this typology here. We agree to Gravina’s observation: BP still licenses null subject, but with context restriction. Our goal here is precisely to investigate which contexts are these. In other words, we will investigate the contexts that allow and favor null subjects in BP, showing that null subjects are less frequent and depend on these contexts to occur. In the next section, we start with verbal morphology and its relation to null subject in BP.

1.1. Verbal morphology

In BP, there are several cases where verbal agreement (AGR) morphology can mark the subject (1st, 2nd, 3rd persons, singular or plural). The relationship between verbal morphology and null subjects can be seen in the sentences below: in (3), verbal morphology marks the subject, i.e., the verb has an exclusive inflectional suffix for the grammatical person; in (4), the verb has syncretic morphology; as a result, the subject cannot be identified by AGR morphology alone in (4), only in (3):

3) Adoro pizza.
Love.1st.person.sg pizza
‘I love pizza’

4) Adora pizza.
Love. $\left. \begin{array}{l} 2^{\text{nd}}.\text{person.sg} \\ 3^{\text{rd}}.\text{person.sg} \\ 1^{\text{st}}.\text{person.pl} \end{array} \right\}$ pizza.

‘You\he\she\it\we love(s) pizza’.

In the past tense, the syncretism is even bigger than in the present, as we show below:

	Present tense:	Past tense:
Eu	ador-o	ador-ava
Você ⁴	ador-a	ador-ava
Ele\ela	ador-a	ador-ava
A gente ⁵	ador-a	ador-ava
Vocês	ador-am	ador-avam
Eles\elas	ador-am	ador-avam

⁴ We also find the pronoun *tu* as a 2nd person singular form. The verb, nevertheless, does not inflect any different in non-monitored speech (see table 1, below).

⁵ We also find the pronoun *nós* as a 1st person plural form. In the present form of the verb, the inflection is different (“*nós ador-ávamos*”); in the past tense, it remains the same (“*nós ador-avam*”) in non-monitored speech.

Until the work of Huang (1984), it seemed clear that the licensing of null subjects in a language would be related to verbal AGR morphology in the following sense: if a language has rich AGR morphology, null subjects could occur (as in Spanish, European Portuguese or Italian); if a language has ‘poor’ verbal AGR morphology, null subjects would not occur (as in English or French) – cf. also Jaeggli; Safir (1989) and Meillet (1918). However, Huang showed that there are languages, such as Chinese and Japanese, for example, that, despite not having rich AGR morphology, license null subjects. Thus, morphology proved to be incapable of explaining the null subject phenomenon, at least independently. As Duarte (1995, p. 2) points out, “the relationship between rich inflection and null subject is no longer exclusive in the processes of licensing and recovery of the null subject’s content.”

In BP, however, there is a relationship between the loss of verbal AGR morphology and the gradual loss of null subjects. In the table below (adapted from DUARTE, 1993, p. 109), we can see, through three periods of time, how the inflectional paradigm of BP has become impoverished:

Table 1: Verbal inflection in BP, present tense

	Paradigm 1	Paradigm 2	Paradigm 3
1st person singular	falo ⁶ (<i>speak+o</i>)	falo	falo
2nd person singular	falas (<i>speak+s</i>) fala (<i>speak+Ø</i>)	fala	fala
3rd person singular	fala (<i>speak+Ø</i>)	fala	fala
1st person plural	falamos (<i>speak+mos</i>)	falamos	(falamos) fala
2nd person plural	falais (<i>speak+is</i>)	falam	falam
3rd person plural	falam (<i>speak+m</i>)	falam	falam

Adapted from DUARTE, 1993, p. 109

In table 1, we see the present tense of the verb ‘falar’ (to speak) in the present tense. Paradigm 3 corresponds to current spoken BP. In this paradigm there are three main forms: 1st p. sing. (*falo*), 2nd p. sing./3rd p. sing./1st p. pl. (*fala*) and 2nd p. pl./3rd p. pl. (*falam*) – and also the less used 1st p. pl. (*falamos*), much less used than ‘fala’. BP’s verbal system is even more syncretic in the past tense (‘pretérito imperfeito’), as we have just shown.

Duarte (1993, p. 110) states that “with a paradigm so impoverished or weakened, nothing is more natural than to expect profound changes in the representation of the pronominal subject.” In view of this shift in the inflection paradigm (from paradigm 1 to 3), morphology is no longer enough a criterion to identify the subjects. Soares; Miller; Hemforth (2019) carried out a quantitative study regarding the effect of verbal agreement on the use of null and pronominal subjects, in relation to the first person in BP. In this study, the authors make a reanalysis of the corpus used by Duarte (1995). According to them,

⁶ The verbal form *falo* is actually the verb stem *fal-* (which is the infinitive form *falar* without the infinitive morpheme *-r* and the thematic vowel *-a-*) with the morpheme *-o* added to its end.

the crucial case for evaluating the effects of the impoverishment of verbal morphology on the choice between overt and null subjects in present-day BP is the first-person singular inflectional marking system. Contrary to second-person markers, which systematically converge with the third person, leading to the precipitous increase in the number of overt subjects [...], first-person subjects differ according to Tense Type [...]. Some tenses have an exclusive marking for first-person singular, others are ambiguous. If verbal inflectional marking is a significant factor in the choice between an overt or a null subject, the Tense Type of the verb should be a significant factor in the choice of null vs. overt subjects for the first-person singular. But it should not interact with the third discourse person singular and plural, since the forms for these are systematically ambiguous with those of the second discourse singular and plural, irrespective of the Tense Type of the verb. The results obtained in this study partially support this prediction (SOARES; MILLER; HEMFORTH, 2019, p. 3586).

Verbal morphology plays a role favoring null subjects in BP, but that is certainly not the only grammatical factor involved in the phenomenon of null subjects. In the next section, we discuss semantic features of the referents of anaphoric subjects and their relation to null and pronominal subjects.

1.2. Semantic features

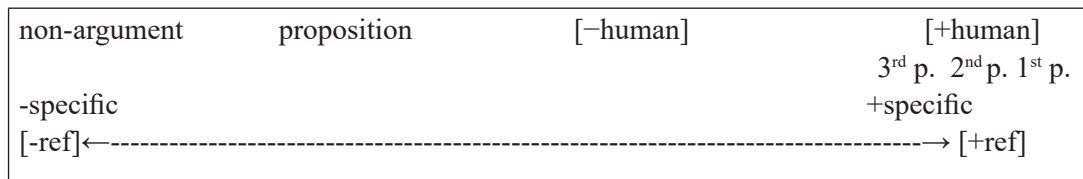
1.2.1 [\pm human] and [\pm specific]

Instigated by the work of Tarallo (1983) – which presented the decrease in the use of null subjects and the increase in the use of pronominal direct objects in BP –, Cyrino; Duarte; Kato (2000) investigated the two phenomena and sought a unified explanation for them. According to them,

If null subjects are licensed by ‘rich’ morphology (TARALDSEN 1978, JAEGGLI; SAFIR 1989, ROBERTS 1993), and if BP’s agreement inflection has been undergoing a change in the direction of a poorer morphology, the decrease of null subjects is justified. But, it is a mystery why the null object should have undergone the opposite trend over time.

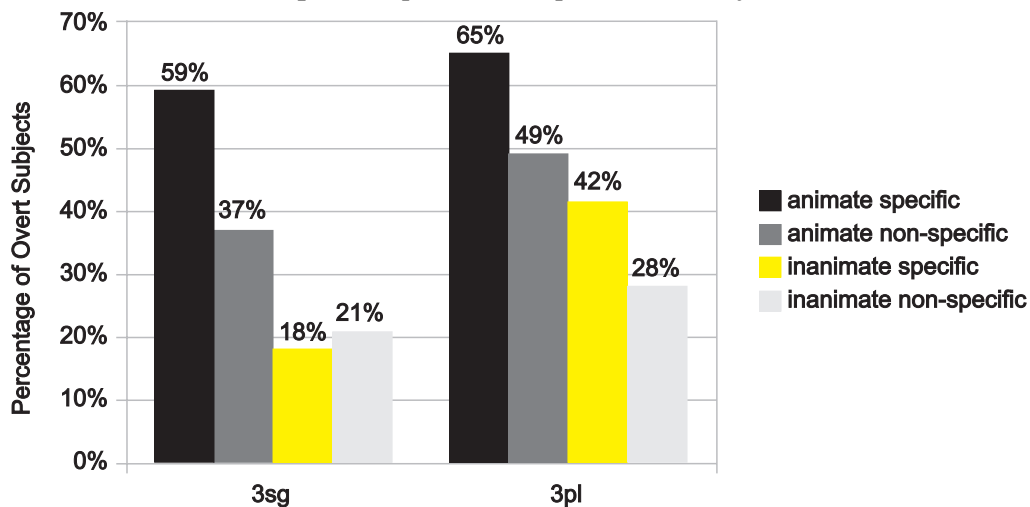
If, on the other hand, null subjects and objects in BP are like those in Chinese (see HUANG, 1984), in that they are not identified by inflection, but rather ‘controlled’ by a co-referent [...], one might expect that, for functional reasons, one or the other, or both, could be null. (CYRINO; DUARTE; KATO, 2000, p. 55).

Cyrino; Duarte; Kato (2000) proposed a referential hierarchy, in which elements at the highest point of the scale would favor pronominal forms, while elements at the lowest point of the scale would favor null forms. The authors present this hierarchy based on the idea that “cross-linguistically, referentiality is highly relevant for pronominalization” (CYRINO; DUARTE; KATO, 2000, p. 58).

Graph 2: Referentiality Hierarchy

Source: adapted from CYRINO; DUARTE; KATO, 2000, p. 59

The farther to the right, the higher the point in the hierarchy. Elements that have the [+human] feature stay at the top, and, therefore, tend to be referred to by pronominal forms. At the opposite end, the referents are preferably referred to by null elements. Cyrino, Duarte; Kato (2000, p. 55) state that “facts found in two independent studies working specifically with the diachronic distribution of null and full pronouns in subject position (Duarte 1993, 1995) and in object position (CYRINO, 1993, 1994[7])” can sustain this analysis – to which we agree. We find difficult, though, to present generalizations based on the analysis considering only the features [\pm human] and [\pm specific]. Soares (2017), for example, analyzes the corpus used by Duarte (1995, the NURC-RJ) in order to investigate the role of [animacy] and [specificity] in null subjects. He then presents the following results:

Graph 3: 3rd person overt pronominal subjects

Source: SOARES, 2017, p. 53

Soares states that

the features Animacy and Specificity take over the purported role of the impoverishment of verbal morphology in BP and explain straightforwardly the imbalanced number of null subjects across discourse persons. However [...] **these two features are not sufficient on their own**” (our emphasis added) (SOARES, 2017, p. 12).

Investigating the same corpus, Soares; Miller; Hemforth, (2020, p. 18) state that

there is a clear tendency: animate and specific antecedents are mostly retrieved by overt subjects while inanimate and non-specific antecedents are preferably recovered by null subjects. The third person singular subject is more frequently overt when its antecedent is animate and specific, followed by the animate non-specific antecedent, and inanimate specific and non-specific antecedents are more or less at the same level; in the third person plural, animate specific overt subjects are close in frequency to other discourse persons, followed by animate non-specific subjects, then by inanimate specific subjects with slightly lower frequency and, at the bottom of the scale, inanimate non-specific subjects.

In other words, we can again verify trends and tendencies, but not definitive generalizations to explain the null subject phenomenon in BP.

1.2.2 Semantic gender of the referent

In addition to Cyrino; Duarte; Kato (2000), other researchers have been working with a different, but related, semantic feature: the semantic gender of the referent. Spinelli (2018), Otero; Spinelli (2019a, b) and Otero; Goldnadel (2020) investigated the relationship of the semantic gender of referents with null subjects in BP. The authors base their hypothesis on Creus; Menuzzi (2004), who demonstrate that natural gender agreement is responsible for 3rd person anaphoric direct objects in BP. BP has a masculine pronoun that refers to male beings and referents ('ele', he), a feminine pronoun that refers to female beings ('ela', she) and an empty category (possibly, little *pro*) that refers to neutral beings and objects (\emptyset , it), a parallel to the English pronominal system, for example. This idea was then, generalized to cover both null object and null subject phenomena in BP.

Semantic gender is not equivalent to grammatical gender, although the two may overlap. In the words of Ayres (2016, p. 22)

The semantic gender feature differs from the grammatical gender feature: the first is "natural", it is about the semantic classification of beings denoted by nouns, which can have natural sex identifiable and recognizable by the speaker; the second refers to the morphosyntactic classification of nouns, which will determine their grammatical agreement relationships. Therefore, it can be said that all nouns in Portuguese have grammatical gender (...). Among the nouns, some will have an inherent semantic gender and others will not. To exemplify: among the nouns *girl*, *cow*, *victim*, and *bottle*, only the first two have inherent semantic gender, but all four have grammatical gender [in Portuguese, all these four nouns are feminine: *a menina*, *a vaca*, *a vítima* and *a garrafa*]. That is, although they are all feminine (grammatical gender), only the first two belong to the class of nouns with the feature [+semantic gender]; the other two nouns ("victim" and "bottle") are marked as [-semantic gender] nouns.

Creus; Menuzzi (2004, p. 16), as we said, propose the semantic gender hypothesis to explain 3rd person anaphoric direct objects, based on the premise that the basic difference between the pronominal forms 'ele/ela' (he/she) and the empty category is that pronouns carry gender features, whereas null forms are not gender-specified. Thus, the choice for an overt pronoun or an empty category, in BP, would be the result of, in the words of the authors (2004, p. 7), a "process of agreement between the antecedent and the anaphoric form." Antecedents with semantic gender would then favor the

use of overt pronouns, since pronouns have gender specification. In contrast, antecedents without semantic gender would favor null objects, since null forms do not have specification for gender. This hypothesis has been pursued with some success (cf. OTHERO; LAZZARI, 2022, for a survey of the latest advances), but it does not have full empirical coverage.

1.3. Avoid V1

Berlinck (1989, 1995), based on diachronic analysis, showed that, in past synchronies, BP had a high use of constructions of the V1 type (VSO, more specifically). That changed over time (cf. also MARTINS; CAVALCANTE; COELHO, 2020), leading to the SVO order in contemporary BP.

In European Portuguese, a null subject language, the first element of the sentence is often a verb (cf. GUILHERME, 2016). In BP, however, there is a great tendency to filling in the initial position of the clause, which is usually the subject position. According to Guilherme (2016, p. 95),

European Portuguese, being a language that maintains characteristics of a null subject language, exhibits a high occurrence of V1 clauses. BP, in turn, seems to have characteristics of a V2 language, as it presents a finite verb in the second position of the sentence, preceded by only one constituent.

Kato; Duarte (2018, pp. 621-2), following Kato (2000)'s suggestion, also point to the fact that

BP rejects the verb in sentence-initial position, filling it with an adjunct or a discursive element. The phenomenon is analyzed as distinct from the syntactic V2 structures, as any category can fill this position: a syntactic adjunct (whether a head or an XP) or a discursive element. Kato's hypothesis is that the constraint here is rhythmic or prosodic.

We will argue here that, in BP, null subjects are allowed when some other constituent occupies the position in the leftmost part of the clause, 'performing the function' of phonetically filling the left periphery of the verb, be it an adjunct or a discourse marker (e.g. *Ontem Ø vi a Maria* [Yesterday (I) saw Maria] should be better/more frequent than *Ø Vi a Maria ontem* [(I) saw Maria yesterday]).

1.4. Discursive Connection

In addition to the generative tradition, there are functionalist studies on null subject in BP. Paredes Silva (1988, 2003), for example, investigated the pronominal subject in BP, bringing together functionalism and Labovian sociolinguistics. From this perspective, she demonstrated that the variation we find between null subjects and overt pronouns has discursive motivations, specially related to topic chain and the 'activation' of the referent in discourse. This idea dates at least to Givón (1979, 1983) and Ariel (1988), i.e., the idea of 'predictability of information': the more predictable a referent (a subject, for example), lesser the need to make it (morphologically\phonetically) explicit. According to Paredes Silva (1998, p. 125),

our conception of the functional use of the pronoun also incorporates Givón's principle of iconicity (...). When applied to the choice between full phrases, strong or weak pronouns and zero anaphora for encoding a third-person referent, the principle (...) predicts that the more predictable an information, the less encoding it receives.

She also states that “the choice of the pronoun is strongly correlated with the non-maintenance of the same referent as a subject” (PAREDES SILVA, 2003, p. 104) – cf. also Bentivoglio (1980) and Lira (1982) on the relation of topic continuity and null subjects in Spanish and Portuguese, respectively.

From a generative point of view, Gravina (2008) has already pointed out that the interpretation of null subjects in BP could depend on an antecedent expressed in the discursive context. According to her, “the empty category in the subject position could have a different nature, with its interpretation guaranteed not by inflection, but by an antecedent expressed in the syntactic, discursive or pragmatic context” (GRAVINA, 2008, p. 21), following Huang (1984)'s suggestion for Chinese.

In the next section, we will show how all these four factors interact in analyzing the null subject phenomenon in BP. We will argue that, together, they can provide an explanation that fully accounts for the phenomenon of the null subject in this language.

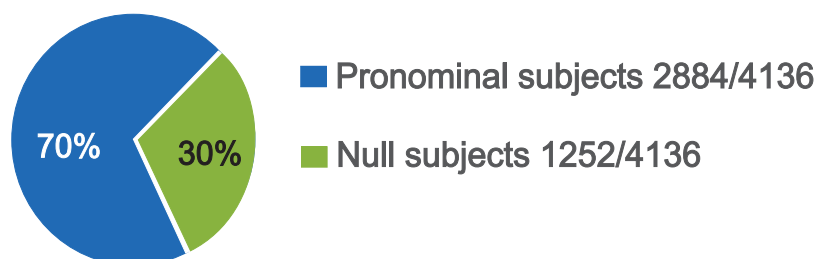
2. Data Analysis

Here we present our results after careful corpus examination. All the data refers to our analysis of the oral language corpus *LínguaPOA* (a corpus composed of sociolinguistic interviews with informants from the city of Porto Alegre, Brazil, cf. Battisti 2019 for information about the corpus and its informants).

2.1 Total data analysis

After analyzing all the interviews, we found 4,136 occurrences of null and pronominal subjects: 2,884 were pronominals and 1,252 were null subjects, as we show in graph 4.

Graph 4: Null and pronominal subjects in the corpus

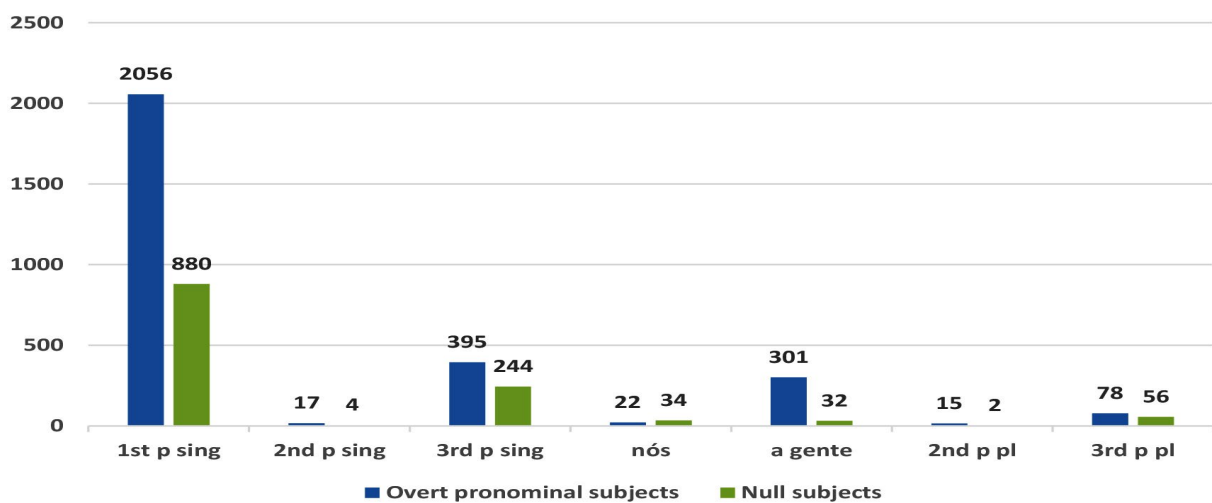


Source: the authors

Null subjects, which, as mentioned above, constituted the majority of subjects in previous periods of BP, are now only residual, 30% of the occurrences. This distribution (~70/~30) has been long found in the literature, at least since the 1990s (cf. BERLINCK; DUARTE; OLIVEIRA, 2015 for a review of the 1990s data). Even though this might lead to think that we are facing a stable phenomenon, as some may suggest (cf. AYRES, 2021, for example), analysis of this same corpus has shown clear indicatives of change in apparent time (cf. OTHERO; LAZZARI, 2022).

With regard to the distribution of null and pronominal subjects and grammatical persons, the vast majority of occurrences were with the 1st person singular, both for null subjects and for pronominal subjects, as shown in the following graph:

Graph 5: Null and pronominal subjects according to grammatical person



Source: the authors

These data are interesting, since they already give a spoiler of the role of morphology: the exponentiation of the subject in verbal morphology seems to be interacting with the null subject phenomenon. For example, notice that BP has two pronouns to express the notion of first person plural: ‘nós’ and ‘a gente’. When we compare these two pronouns, we find more null subjects precisely with ‘nós’, since in this case the verb exhibits exclusive first person plural morphology (60.8%). On the other hand, the pronoun ‘a gente’, which does not have AGR morphology, is almost always realized as a pronoun (90.3%).

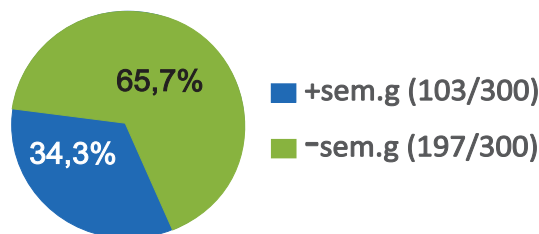
2.2 Analysis of null subject data

As we have seen, the literature has been reporting that overt subjects are the default case in the Brazilian Portuguese. Null subjects are a ‘marked’, less frequent strategy. After investigating the four factors presented in the previous sections, we analyzed all occurrences of null subjects in the corpus to check if they can all be explained – or justified – by one or more of these factors that are known for favoring null subjects.

Our main hypothesis here is that there is no ‘one’ licensing context or factor for null subjects in BP. Instead there are four: **semantic gender of the referent** (referents [-gs] favor null subjects), **topic-chain** or discourse saliency (salient referents in discourse tend to be null subjects), **linear V2 pattern** (null subjects are not favored in V1 clauses) and **verbal morphology** (overt AGR morphology favors null subjects). If this hypothesis is right, these four factors should be necessary and sufficient to ‘explain’ all null subject occurrences in any given corpus of contemporary spoken BP.

Let’s start with the first factor mentioned in the previous paragraph: the semantic gender. That applies only to 3rd personal subjects, since these are the pronouns that have morphological forms for masculine (*ele\eles*, he\they.masc) and feminine (*ela\elas*, she\they.fem). They compete with the null, underspecified form for gender (\emptyset). We found 300 occurrences of 3rd person null subjects, 65.7% of them were coreferent to [-sem.g] referents. That means the other 34.3% occurrences were ‘unexpected’, considering only the semantic gender hypothesis.

Graph 6: Null subjects and their referents [\pm sem.g]



Source: the authors

Examples:

- 5) *A internet mesmo eu uso praticamente vinte e quatro horas por dia... [O meu celular] tá sempre conectado, quando eu tô em casa, \emptyset tá sempre no wifi, quando eu tô na rua, \emptyset tá sempre nos dados.* (LínguaPOA – Interview 06)⁷
- 6) *Eu vejo bastante jovens trabalhando desde/ de cedo, eu tenho [uma prima] que tá com dezesseis anos e \emptyset tá fazendo o jovem aprendiz.* (LínguaPOA – Interview 92)⁸

In (5), we see two null subjects referring to [O meu celular] (‘my cellphone’), as expected, considering only the semantic gender factor. In (6), the null subject refers to [uma prima] (‘a female cousin’), a referent with semantic (apparent) gender. In this case, the null subject is not licensed by the semantic gender of the referent; it is licensed, though, by the discursive context: it appears in a coordinate clause (after *e*, ‘and’), a universal context favoring null subjects.

⁷ Roughly translated as “The internet itself, I use it practically 24h per day... [My cellphone] is always connected when I’m home, it is always with the wi-fi on when I’m on the street, it is always in 4G mode.”

⁸ Roughly translated as “I see lots of young people working sin/since an early age, I have [a cousin] who is 16 and she is in a young trainee program.”

The second factor, referent saliency in discourse (i.e. referents participating in a topic chain), could not cover all the cases of null subjects either: 54.3% of the null subjects were in a context of topic chain, but 45.7% were not. Here we analyzed all the 1,252 occurrences of null subjects in the corpus.

Graph 7: Null subjects and discourse saliency

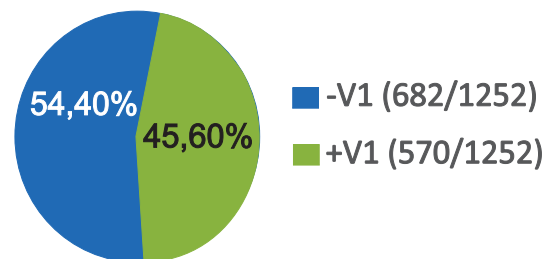


Source: the authors

The third factor, Avoid V1 linear order, could not explain all the occurrences of null subjects we found as well: from the 1,252 cases of null subjects, only 682 (54.4%) did not occur in V1 clauses. That is to say that examples like (7) were relatively frequent:

- (7) *E tu gosta de cozinha(r)?*
 Ø *Gosto muito.* (LínguaPOA – Interview 30)⁹

Graph 8: Null subjects and the *V1 effect



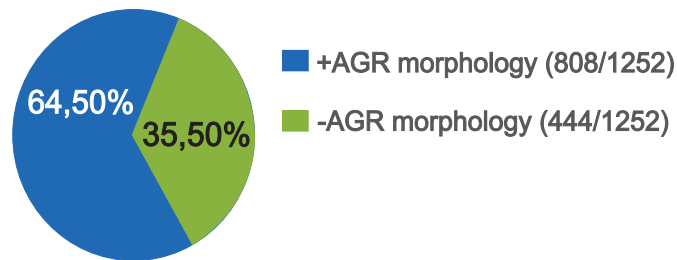
Source: the authors

Finally, when we analyzed the fourth factor, verbal inflectional AGR morphology, we found a slightly higher influence on null subjects: 64.5% of the null subjects found in the corpus occurred with verbs that have distinctive AGR morphology.

- (8) *E tu gosta de cozinha(r)?*
 Ø *Gosto muito.* (LínguaPOA – Interview 30)

The verb ‘gosto’ (I like) presents the ending -o (*gost-o*), marking 1st person singular exclusively.

⁹ Roughly translated as: “And you, do you like cookin’? I like it a lot.”

Graph 9: Null subjects and verbal AGR morphology

Source: the authors

In sum, we can see that none of the factors, in isolation, can be *the* factor licensing null subjects in BP. In the next section, we will show an analysis considering combinations of two factors, in order to check whether these combinations are able to explain all occurrences of null subjects in the corpus.

2.3. Two factors combined

Let's analyze all the possible combinations of the factors, grouping them together, in pairs, namely:

- i) semantic gender and discourse saliency (sem.g & ds, cf. table 2);
- ii) semantic gender and *V1 (sem.g & *V1, see table 3);
- iii) semantic gender and verbal morphology (sem.g & morph, see table 4);
- iv) discourse saliency and *V1 (ds & *V1, cf. table 5);
- v) discourse saliency and verbal morphology (ds & morph, see table 6);
- vi) *V1 linear pattern and morphology (*V1 & morph, cf. table 7).

In order to make it clearer for the reader, we will use the symbols ♡ when the context favors null subject and ☹ when it does not favor null subject. In the table below, we present the combination of semantic gender and discourse saliency:

Table 2: Semantic gender and discourse saliency of the referent

Semantic gender	Discourse saliency	Null subjects
[-sem.g] ♡	[+ds] ♡	93/300 (31%)
[-sem.g] ♡	[-ds] ☹	104/300 (34,7%)
[+sem.g] ☹	[+ds] ♡	72/300 (24%)
One or two favoring factors licensing the null subject		269/300 (89.7%)
[+sem.g] ☹	[-ds] ☹	31/300 (10,3%)









Source: the authors

With the combination that has the two factors that do not favor the null subject (the combination [+sem.g] and [-ds]), we would expect not to find any null subjects. However, as we can see in table 2, there are 31 occurrences (10.3%) of null subjects in this context; the other 269 occurrences (89.7%) appeared in a context with at least one licensing factor (either -sem.g or +ds, or both).

At this point, we can already see that with just these two factors it is not possible to explain the data of null subjects found, nor to make predictions about the data of null subjects that might be found in BP.

The second pair of factors we have analyzed is the semantic gender with linear order. The distribution of null subjects according to these traits is presented in the table below:

Table 3: Semantic gender of the referent and & V1 linear order.









Semantic gender	linear order	Null subjects
[-sem.g] 	[-V1] 	118/300 (39,3%)
[-sem.g] 	[+V1] 	79/300 (26,3%)
[+sem.g] 	[-V1] 	58/300 (19,3%)
One or two favoring factors licensing the null subject		255/300 (85%)
[+sem.g] 	[+V1] 	45/300 (15%)

Source: the authors

Again, we should expect that ([+sem.g]) referents and clauses where the verb is in the initial position (+V1) would not allow null subjects. However, we found 45 occurrences of null subjects in this scenario. Once again, it is not possible to explain the occurrences of null subjects found in the corpus.

The third possible combination of factors is the combination of semantic gender and verbal AGR morphology. The results found are in table 4:

Table 4: Semantic gender of the referent and AGR morphology

Semantic gender	AGR morphology	Null subjects
[-sem.g] 	[+morph] 	0/300 (0%)
[-sem.g] 	[-morph] 	197/300 (65.7%)
[+sem.g] 	[+morph] 	0/300 (0%)
One or two favoring factors licensing the null subject		197/300 (65.7%)
[+sem.g] 	[-morph] 	103/300 (34.3%)









Source: the authors

Here we cannot find any occurrence in a [+morph] context, since we are analyzing only 3rd person subjects and the verbs in 3rd person always present poor morphology (cf. discussion in section 2.1). Most occurrences of null subjects are in a context where they refer to [-gs] referents (as we have seen in 3.2). Nonetheless, we have found more than a third of the null subjects in the ‘unexpected’ context of [+sem.g] and [-morph].

The data presented in the tables above contained only occurrences of 3rd person null subjects, since we were considering the semantic gender feature. In the next tables, we will consider the combination of the other factors, and they are relevant for all three discourse persons.

In table 5, we present the combination of discourse saliency and V1 linear pattern.

Table 5: Discourse saliency of the referent and V1 linear order









Discourse saliency	V1 linear order	Null subjects
[+ds] 	[-V1] 	300/1,252 (23.9%)
[+ds] 	[+V1] 	380/1,252 (30.3%)
[-ds] 	[-V1] 	382/1,252 (30.5%)
One or two favoring factors licensing the null subject		1,062/1,525 (84,9%)
[-ds] 	[+V1] 	190/1,252 (15.1%)

Source: the authors

We still find a great number of null subjects in ‘unexpected’ scenarios, in this case, in [-ds] and [+V1] contexts. That clearly shows that these two factors are far from explaining the contexts where null subjects can occur in BP.

The penultimate combination of two factors is the one that brings together discourse saliency and morphology.

Table 6: Discourse saliency of the referent and verbal AGR morphology

Discourse saliency	AGR morphology	Null subjects
[+ds] 	[+morph] 	426/1,252 (34%)
[+ds] 	[-morph] 	254/1,252 (20.2%)
[-ds] 	[+morph] 	382/1,252 (30.5%)
One or two favoring factors licensing the null subject		1,062/1,525 (84,9%)
[-ds] 	[-morph] 	190/1,252 (15.1%)

Source: the authors

The majority of occurrences were in a context licensed by both factors, i.e., the null subject was in a context where it could be retrieved both by verbal morphology and discourse saliency, as we show in the example below:









- (9) *E tu trabalha? Estuda? Que tu faz?*
 Ø *Trabalho aqui, ahm, como auxiliar administrativo* (LínguaPOA – Interview 60)¹⁰.

Nevertheless, 15.1% of the cases could not be explained by neither of these factors, i.e. null subjects that occurred in [-ds] and [-morph] contexts.

Finally, the last combination of two factors is the one that unites linear V1 order and verbal AGR morphology.

¹⁰ Roughly translated as: “And you, do you work? Study? What do you do? I work here, hum, as an administrative assistant.”

Table 7: V1 and verbal AGR morphology

V1 linear order	AGR morphology	Null subjects
[-V1] 	[+morph] 	415/1,252 (33,1%)
[-V1] 	[-morph] 	267/1,252 (21,3%)
[+V1] 	[+morph] 	393/1,252 (31,3%)
[+V1] 	[-morph] 	177/1,252 (14,1%)

Source: the authors

























Again, the majority of null subjects appeared in a context licensed by both factors, i.e. the null subject was in a context where it could be retrieved by verbal morphology and was not in a clause where the verb was the first and leftmost element of the clause. Nonetheless, that was, again, just roughly a third of the cases found. The combination of two factors is clearly not enough to cover all null subjects in the corpus.

Our next step was, then, to check whether three factors could account for all null subjects occurrences, as we will see in the next section.

2.4. Three Null Subject Licensing Factors

In this section we present sets of three factors and see if these sets can explain all occurrences of null subjects in the corpus. The combinations of factors are (i) semantic gender, discourse saliency and V1 linear order (table 8); (ii) semantic gender, V1 linear order and AGR morphology (table 9); (iii) semantic gender, discourse saliency and AGR morphology (table 10); and (iv) AGR morphology, V1 linear order and discourse saliency (table 11).

Table 8: Semantic gender of the referent, discourse saliency and V1 linear order

Semantic Gender	Discourse Saliency	V1 Linear Order	Null Subjects
[+sem.g] 	[+ds] 	[-V1] 	32/300 (10,6%)
[+sem.g] 	[+ds] 	[+V1] 	40/300 (13,3%)
[+sem.g] 	[-ds] 	[-V1] 	26/300 (8,6%)
[+sem.g] 	[-ds] 	[+V1] 	5/300 (1,6%)
[-sem.g] 	[+ds] 	[-V1] 	49/300 (16,3%)
[-sem.g] 	[+ds] 	[+V1] 	44/300 (14,6%)
[-sem.g] 	[-ds] 	[-V1] 	69/300 (23%)
[-sem.g] 	[-ds] 	[+V1] 	35/300 (11,6%)

Source: the authors

Once again, we could not cover all cases of null subjects, considering these three licensing factors. We found null subjects (only 5 occurrences though) in the unexpected scenario of [+sem.g], [-ds] and [+V1]. That means we could understand and ‘explain’ the other 295 occurrences of null subjects, but bear in mind we are only analyzing 3rd person subjects here.

Next we analyze semantic gender, V1 linear order and AGR morphology.

Table 9: Semantic gender of the referent, V1 linear order and AGR morphology

Semantic Gender	V1 linear order	AGR morphology	Null subjects
[+sem.g] ☞	[-V1] ☞	[+morph] ☞	0/300 (0%)
[+sem.g] ☞	[-V1] ☞	[-morph] ☞	58/300 (19,3%)
[+sem.g] ☞	[+V1] ☞	[+morph] ☞	0/300 (0%)
[+sem.g] ☞	[+V1] ☞	[-morph] ☞	45/300 (15%)
[-sem.g] ☞	[-V1] ☞	[+morph] ☞	0/300 (0%)
[-sem.g] ☞	[-V1] ☞	[-morph] ☞	118/300 (39,3%)
[-sem.g] ☞	[+V1] ☞	[+morph] ☞	0/300 (0%)
[-sem.g] ☞	[+V1] ☞	[-morph] ☞	79/300 (26,3)













Source: the authors

Again, we could not explain the phenomenon entirely, since we found 45 occurrences of null subjects in a very disfavoring context: [+sem.g] referents, [+V1] linear order and verbal forms with poor morphology [-morph]. Comparing Tables 9 and 8, it is clear the discourse saliency is a factor to be considered.

The penultimate combination concerns the semantic gender of the referent, the discourse saliency of the referent and AGR morphology. The occurrences of null subjects were distributed as follows:

Table 10: Semantic gender of the referent, discourse saliency and AGR morphology

Semantic gender	Discourse saliency	AGR Morphology	Null subjects
[+sem.g] ☞	[+ds] ☞	[+morph] ☞	0/300 (0%)
[+sem.g] ☞	[+ds] ☞	[-morph] ☞	72/300 (24%)
[+sem.g] ☞	[-ds] ☞	[+morph] ☞	0/300 (0%)
[+sem.g] ☞	[-ds] ☞	[-morph] ☞	31/300 (10,3%)

























[-sem.g] 	[+ds] 	[+morph] 	0/300 (0%)
[-sem.g] 	[+ds] 	[-morph] 	93/300 (31%)
[-sem.g] 	[-ds] 	[+morph] 	0/300 (0%)
[-sem.g] 	[-ds] 	[-morph] 	104/300 (34,6%)

Source: the authors

Here we found 31 occurrences (10,3%) of null subjects appearing in the unexpected context of [+sem.g], [-ds] and [-morph]. So we still could not explain all the occurrences of null subjects in the corpus.

The last combination of three factors involves discourse saliency, V1 linear order and AGR morphology. Here, we analyzed all occurrences of null subjects in the corpus.

Table 11: Discourse saliency, V1 pattern and AGR morphology

Discourse saliency	V1 linear order	AGR Morphology	Null subjects
[+ds] 	[-V1] 	[+morph] 	174/1,252 (13,8%)
[+ds] 	[-V1] 	[-morph] 	126/1,252 (10%)
[+ds] 	[+V1] 	[+morph] 	252/1,252 (20,1%)
[+ds] 	[+V1] 	[-morph] 	128/1,252 (10,2%)
[-ds] 	[-V1] 	[+morph] 	241/1,252 (19,2%)
[-ds] 	[-V1] 	[-morph] 	141/1,252 (11,2%)
[-ds] 	[+V1] 	[+morph] 	141/1,252 (11,2%)
[-ds] 	[+V1] 	[-morph] 	49/1,252 (3,9%)

Source: the authors

We are certainly narrowing down the cases of null subjects. If we assume that three of the four factors commonly assumed by the literature are acting together to license null subjects in BP, we can already explain lots of cases we found in the corpus. That's pretty good, but we are not telling the whole story yet. That is why we advocate for a broader analysis: in the next section we will demonstrate that all four factors are needed to fully understand the null subject phenomenon in BP.

2.4.1. Four Null Subject Licensing Factors

As we have seen so far, taking in consideration only one licensing factor in isolation did not lead to any generalization. As we increased the number of factors to two or three in the analyzed sets, we found better results when analyzing the occurrences of null subjects in the corpus. However, even taking into account three factors, we still found 49 cases of null subjects that were unexpected (see

table 11). Since we aim to reach an explanation of the totality of the null subjects in the corpus (i.e., the 1,252 occurrences found), we will analyze the data using the four factors presented here. The combinations of four null subject licensing factors that occurred were shown in the next table¹¹.

Table 12: Semantic gender of the referent, discourse salience, V1 linear order and AGR morphology

Semantic gender	Discourse salience	V1 linear order	AGR Morphology	Null subjects
[+sem.g] 🖱	[+ds] 🖱	[-V1] 🖱	[+morph] 🖱	🖱 174/1,252 (13.8%)
[+sem.g] 🖱	[+ds] 🖱	[-V1] 🖱	[-morph] 🖱	🖱 75/1,252 (5.9%)
[+sem.g] 🖱	[+ds] 🖱	[+V1] 🖱	[+morph] 🖱	🖱 252/1,252 (20,1%)
[+sem.g] 🖱	[+ds] 🖱	[+V1] 🖱	[-morph] 🖱	🖱 92/1,252 (6.5%)
[+sem.g] 🖱	[-ds] 🖱	[-V1] 🖱	[+morph] 🖱	🖱 241/1,252 (19.2%)
[+sem.g] 🖱	[-ds] 🖱	[-V1] 🖱	[-morph] 🖱	🖱 72/1,252 (5,7%)
[+sem.g] 🖱	[-ds] 🖱	[+V1] 🖱	[+morph] 🖱	🖱 141/1,252 (11.2%)
[+sem.g] 🖱	[-ds] 🖱	[+V1] 🖱	[-morph] 🖱	🖱 3/1,252 (1%)
[-sem.g] 🖱	[+ds] 🖱	[-V1] 🖱	[+morph] 🖱	🖱 0/1,252 (0%)
[-sem.g] 🖱	[+ds] 🖱	[-V1] 🖱	[-morph] 🖱	🖱 51/1,252 (4%)
[-sem.g] 🖱	[+ds] 🖱	[+V1] 🖱	[+morph] 🖱	🖱 0/1,252 (0%)
[-sem.g] 🖱	[+ds] 🖱	[+V1] 🖱	[-morph] 🖱	🖱 46/1,252 (3.6%)
[-sem.g] 🖱	[-ds] 🖱	[-V1] 🖱	[+morph] 🖱	🖱 0/1,252 (0%)
[-sem.g] 🖱	[-ds] 🖱	[-V1] 🖱	[-morph] 🖱	🖱 69/1,252 (5.5%)
[-sem.g] 🖱	[-ds] 🖱	[+V1] 🖱	[+morph] 🖱	🖱 0/1,252 (0%)
[-sem.g] 🖱	[-ds] 🖱	[+V1] 🖱	[-morph] 🖱	🖱 36/1,252 (2.8%)

Source: the authors

With the combination of these four factors, we achieved a better result: we account for 99.8% (or 1,249 occurrences) of all the 1,252 null subjects found in the corpus. Among all the occurrences of null subjects, only 3 were, at a first sight, unexpected. That is an encouraging result. It means that these four factors play a role for defining the contexts for null subjects in Brazilian Portuguese.

Final remarks

As we said at the beginning, the null subject phenomenon has already been extensively investigated in BP. Much has been discovered, but some questions remain, whose answers are open.

¹¹ We have incorporated all the occurrences in the table, 1st, 2nd and 3rd persons, singular and plural. The semantic gender trait is only valid for 3rd persons.

One of these, which motivated our research, was to narrow down the contexts that allow null subjects in BP. Unlike most of the literature on the topic, we proposed here that four factors play a role in determining the optimal contexts for null subjects. They are both necessary and sufficient to understand and explain all null subject occurrences in any given corpus of contemporary spoken BP. We have presented here a multifactorial analysis, uniting the factors semantic gender, verbal morphology, optimal discursive connection and linear order to account for the null subject phenomenon in BP. We have analyzed the data taking only one factor in consideration, then two, three and finally all the four factors. As we increased the number of factors, the more consistent our analysis proved to be. With one factor alone, we could not point out robust tendencies, but as we started to analyze the factors as a set, we obtained better results, until we have arrived at the occurrences analyzed from the set of four factors, which, as we have seen, explains 99.8% of data, leading us to a better understanding of the null subject phenomenon in BP, at least when it comes to identifying its licensing contexts.

References

- ARIEL, Mira. Referring and accessibility. *Journal of linguistics*, v. 24, n. 1, pp. 65-87, 1988.
- AYRES, M. R. *Aspectos condicionadores do objeto nulo e do pronome pleno em português brasileiro: uma análise da fala infantil*. Dissertação de mestrado, Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS, 2016.
- AYRES, M. R. *Contextos licenciadores de sujeitos nulos em português brasileiro*. 123 f. PhD dissertation. Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021.
- BATTISTI, E. *O acervo de entrevistas sociolinguísticas LínguaPOA: constituição, possibilidades e desafios*. Comunicação oral proferida no I Fórum Internacional de Sociolinguística, UFRJ, 26 nov. 2019.
- BENTIVOGLIO, P. A. *Why ‘canto’ and not ‘yo canto’? The problem of first person subject pronoun in spoken Venezuelan Spanish*. 72 f. M.A. thesis. University of California, 1980.
- BERLINCK, R. A. A construção V SN no português do Brasil: uma visão diacrônica do fenômeno da ordem. In: TARALLO, F. (org.) *Fotografias Sociolinguísticas*. Campinas: Pontes, 1989. pp. 95-112,
- BERLINCK, R. A. As posposições do sujeito em português. In: 1º Encontro do CelSul (Círculo de Estudos Linguísticos do Sul), 1997, Florianópolis. *Anais do 1º Encontro do CelSul*, 1995. v. 2. pp. 507-19.
- BERLINCK, R. A.; DUARTE, M. E. L.; OLIVEIRA, M. Predicação. In: KATO, M.; NASCIMENTO, M. (org.) *A Construção da Sentença: Gramática do Português Culto Falado no Brasil*. São Paulo: Contexto, 2015.
- BIBERAUER, T. Semi pro-drop languages, expletives and expletive pro reconsidered. In: BERLINCK, R. A. et al. (org.) *Parametric Variation: null subjects in Minimalist theory*. Cambridge: Cambridge University Press. pp. 153-99, 2010.
- BIBERAUER, T. et al. (org.) *Parametric Variation: Null Subjects in Minimalist Theory*, Cambridge: CUP, 2010.

CHOMSKY, N. *Lectures on Government and Binding*. Dordrecht: Foris, 1981.

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*, v. 14, n. 4, pp. 2606-17, 2017.

CREUS, S.; MENUZZI, S. O papel do gênero na alternância entre objeto nulo e pronome pleno em português brasileiro. *Revista da ABRALIN*, Florianópolis, v. 3, n. 1-2, pp. 149-76, 2004.

CYRINO, S. M. L.; DUARTE, M. E.; KATO, M. A. Visible subjects and invisible clitics in Brazilian Portuguese. In: KATO, M. A.; NEGRÃO, E. V. (org.) *Brazilian Portuguese and the Null Subject Parameter*. Frankfurt: Vervuert-Iberoamericana, pp. 55-104, 2000.

CYRINO, S. M. *O objeto nulo no português do Brasil: um estudo sintático-diacrônico*. Londrina: Ed. da Universidade Estadual de Londrina, 1997.

CYRINO, S. M. 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.

DUARTE, I.; FIGUEIREDO SILVA, M. C. The Null Subject Parameter and the Structure of the Sentence in European and Brazilian Portuguese. In: WETZELS, L.; MENUZZI, S.; COSTA, J. *The Handbook of Portuguese Linguistics*. West Sussex: John Wiley & Sons, 2016. pp. 234-53.

DUARTE, M. E. L. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. In: ROBERTS, I.; KATO, M. *Português brasileiro: uma viagem diacrônica: homenagem a Fernando Tarallo*. São Paulo: Contexto, 2018. pp. 83-104.

DUARTE, M. E. L. *A perda do princípio “evite pronome” no português brasileiro*. 151 f. PhD dissertation. Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1995.

DUARTE, M. E. L. A remarcação em curso no valor do parâmetro do sujeito nulo. *Cuadernos de la Alfal*, v. 12, pp. 71-99, 2020.

DUARTE, M. E. L. 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. pp. 107-28.

DUARTE, M. E. L. *O sujeito em peças de teatro (1833-1992): estudos diacrônicos*. São Paulo: Parábola Editorial, 2012.

DUARTE, M. E. L. ReVEL na Escola: Sobre pronomes pessoais na fala e na escrita. *ReVEL*, v. 16, pp. 1-12, 2018.

DUARTE, M. E. L.; MARINS, J. Brazilian Portuguese: a ‘partial’ null subject language. *Cadernos de Estudos Linguísticos*, v. 63, 2021.

GRAVINA, A. P. *A Natureza do Sujeito Nulo na Diacronia do PB: estudo de um corpus mineiro (1845 a 1950)*. 146 f. M.A. thesis. Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2008.

- GRAVINA, A. P. Diacronia e sujeito nulo no português brasileiro: um estudo comparativo. *Filol. Linguíst. Port.*, São Paulo, v. 16, n. spe, pp. 199-231, dez. 2014.
- GUILHERME, M. R. C. Português Brasileiro: uma língua V2? *Interfacis*, Belo Horizonte, v. 2, n. 1, pp. 82-97, 2016.
- HOLMBERG, A. Is There a Little Pro? Evidence from Finnish. *Linguistic Inquiry* 36, pp. 533-64, 2005.
- HOLMBERG, A.; NAYUDU, A.; SHEEHAN, M. Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish, and Marathi, *Studia Linguistica*, v. 63, n. 1, pp. 59-97, 2009.
- HUANG, C.-T. J. On the distribution and reference of empty pronouns. *Linguistic Inquiry*, v. 15, pp. 531-74, 1984.
- JAEGGLI, O.; SAFIR, K. *The Null Subject Parameter*, Dordrecht: Springer, 1989.
- KATO, M. A. *Determinantes prosódicos em mudança sintática*. Abralin ao vivo, 2020. [<https://www.youtube.com/watch?v=t3BLRPloZJI&t=2925s>]
- KATO, M. A. Strong and weak pronominals in the null subject Parameter. *Probus*, Berlin, v. 11, n. 1, pp. 1-37, 1999.
- KATO, M. A. The partial pro-drop nature and the restricted VS order in Brazilian Portuguese. In: KATO, M. A. (org.). *Brazilian Portuguese and the Null Subject Parameter*, Vervuert- Iberoamericana, Frankfurt, Germany, 2000. pp. 223-58.
- KATO, M. A.; DUARTE, M. E. L. Pre-verbal position in BP: a reinterpretation of “avoid pronoun principle”. *Diadorim*, v. 20 n. esp., pp. 610-26, 2018.
- LIRA, S. A. *Nominal, Pronominal and Zero Subject in Brazilian Portuguese*. PhD Dissertation, University of Pennsylvania, 1982.
- MARTINS, M. A. R.; CAVALCANTE, S.; COELHO, I. L. Ordem do sujeito e colocação de clíticos na escrita brasileira dos séculos XIX e XX. *Cadernos de Estudos Linguísticos*, v. 62, pp. 1-22, 2020.
- MEILLET, A. Convergences des développements linguistiques. *Revue Philosophique de la France et de l'Étranger*, v. 85, pp. 97-110, 1918.
- MONTEIRO, J. L. *Pronomes pessoais: subsídios para uma gramática do português do Brasil*. Fortaleza: Edições UFC, 1994.
- OTHERO, G. A.; GOLDNADEL, M. Omissão de sujeito pronominal anafórico e as construções de dupla negação. *Cadernos de Estudos Linguísticos* (Unicamp), v. 62, pp. 1-21, 2020.
- OTHERO, G. A.; LAZZARI, M. Null subjects and null objects in Brazilian Portuguese: correlations and change. *Revista de Estudos da Linguagem*, v. 30, n. 4, 2022.
- OTHERO, G. A.; SPINELLI, A. C. Sujeito expresso e nulo no começo do séc. XXI (e sua relação com o objeto nulo em PB). *Domínios de Linguagem*. Uberlândia, v. 13, n. 1, 2019a.
- 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. *Caderno de Estudos Linguísticos*, v. 61, n. 1, 2019b.

- PAREDES SILVA, V. L. *Cartas cariocas. A variação do sujeito na escrita informal*. Phd dissertation, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.
- PAREDES SILVA, V. L. Motivações funcionais no uso do sujeito pronominal: uma análise em tempo real. In: PAIVA, M. C.; DUARTE, M. E. L. *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa, 2003. pp. 97-114.
- PERLMUTTER, D. *Deep and Surface Structure Constraints in Syntax*. New York: Holt, Rinehart and Winston, 1971.
- ROBERTS, I. Posfácio. O português brasileiro no contexto das línguas românicas. In: ROBERTS, I.; KATO, M. A. (orgs.) *Português brasileiro: uma viagem diacrônica*. Campinas: Ed. da Unicamp, 1993. pp. 409-25.
- RODRIGUES C. N. *Effects of Loss of Morphology in Partial pro-drop Languages*. PhD dissertation, University of Maryland, Maryland, 2004.
- SOARES, E. C. *Anaphors in discourse: anaphoric subjects in Brazilian Portuguese*. PhD Thesis. 482 f. University Sorbonne Paris City, Paris, France, 2017.
- 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*, v. 36, pp. 1-38, 2020.
- 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, pp. 3579-600, 2019.
- SPINELLI, A. C. *Pronomes e sua ausência: por um tratamento unificado da omissão de sujeitos e objetos diretos pronominais de 3ª pessoa*. 78 f. M. A. thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.
- TARALDSEN, K. T. On the NIC, vacuous application, and the that-trace filter. *Indiana University Linguistics Club*, 1978.
- TARALLO, F. *Relativization strategies in Brazilian Portuguese*. Tese de Doutorado. University of Pennsylvania, Pennsylvania, 1983.
- WIDERA, C.; KAISER, G. A. Asterix e os pronomes sujeitos: uma análise contrastiva do emprego dos pronomes sujeitos no português europeu e brasileiro. In: *I Congresso Internacional em Variação Linguística nas Línguas Românicas*, pp. 143-65, 2019.