

**THE EFFECT
OF VERBAL AGREEMENT
MARKING ON THE USE OF NULL AND
OVERT SUBJECTS –
A QUANTITATIVE STUDY
OF FIRST-PERSON
SINGULAR IN BRAZILIAN
PORTUGUESE**

O EFEITO DA MARCAÇÃO DE CONCORDÂNCIA VERBAL NO USO DE SUJEITOS NULOS E PRONOMINAIS - UM ESTUDO QUANTITATIVO DA PRIMEIRA PESSOA SINGULAR NO PORTUGUÊS BRASILEIRO

EL EFECTO DE LAS MARCAS VERBALES DE ACUERDO EN EL USO DE SUJETOS NULOS Y PRONOMINALES: UN ESTUDIO CUANTITATIVO DE LA PRIMERA PERSONA SINGULAR EN PORTUGUÊS BRASILEÑO

Eduardo Correa Soares*

Universidade Federal de Santa Catarina

Philip Miller**

Université Paris Diderot

Barbara Hemforth***

Université Paris Diderot/Centre National de la Recherche Scientifique

* Federal University of Santa Catarina (UFSC) – Graduate School of Linguistics (PPGL) – Laboratory of Language and Cognitive Processes; Coordination Program of the Improvement of Higher Level Personnel (CAPES). soares_ec@yahoo.com.br.

** University of Paris VII Denis Diderot – CLILLAC-ARP. philip.miller@univ-paris-diderot.fr.

*** University of Paris VII Denis Diderot – LLF; French National Center for Scientific Research (CNRS). barbara.hemforth@linguist.univ-paris-diderot.fr.

ABSTRACT: In a corpus study and two acceptability experiments, we investigated whether there is a preference for null or overt subjects with ambiguous (syncretic) and exclusive (unambiguous) verbs in Brazilian Portuguese (BP). Previous literature has proposed that BP (i) is deactivating the “Avoid Pronoun Principle” and (ii) is a partially pro-drop language, whose morphosyntactic contexts for null subjects are restricted. The corpus study and the acceptability experiments show that there is an effect of person syncretism in the verbal paradigm on the relative frequency of null subjects and on the acceptability of sentences with null subjects in the first-person singular: there is a tendency to avoid ambiguity due to null subjects with ambiguous verb forms, but only in contexts with competing antecedents. These results are analyzed in the light of a general theory of anaphora resolution (ARIEL, 1990, etc.), as resulting from a calculation taking into account the accessibility of potential antecedents and the cost of the anaphoric form (ALMOR, 1996).

KEYWORDS: Null Subject. Verbal paradigm. Ambiguity. Anaphora resolution.

RESUMO: Em um estudo de corpus e em dois experimentos de aceitabilidade, investigamos se existe uma preferência por sujeitos nulos ou pronominais com verbos ambíguos (sincréticos) ou exclusivos (não-ambíguos) no Português Brasileiro (PB). A literatura propõe que o PB (i) está desativando o “Princípio Evitar Pronome” e (ii) é uma língua parcialmente “pro-drop”, cujos contextos morfossintáticos para sujeitos nulos são restritos. Nosso estudo de corpus e os experimentos de aceitabilidade mostram que há um efeito do sincretismo de pessoas no paradigma verbal sobre a frequência relativa de sujeitos nulos e sobre a aceitabilidade de sentenças com sujeitos nulos na primeira pessoa do singular: há uma tendência a evitar ambigüidade para sujeitos nulos com formas verbais ambíguas, mas apenas em contextos em que há competição de antecedentes. Esses resultados são analisados à luz de uma teoria geral da resolução da anáfora (ARIEL, 1990, etc.), como resultado de um cálculo levando em conta a acessibilidade dos antecedentes potenciais e o custo da forma anafórica (ALMOR, 1996).

PALAVRAS-CHAVE: Sujeitos nulos. Paradigma verbal. Ambigüidade. Resolução anafórica.

RESUMEN: En un estudio con base en corpus y dos experimentos de aceptabilidad, investigamos si existe una preferencia por sujetos nulos o pronominales con verbos ambiguos (sincréticos) o exclusivos (no ambiguos) en portugués brasileño (PB). En los estudios del área, se propuso (i) que se está desactivando en PB el “Principio Evite Pronombres” y (ii) que el PB es una lengua parcialmente “pro-drop”, cuyos contextos morfosintáticos para sujetos nulos se han restringido. Nuestro estudio con base en corpus y en los experimentos de aceptabilidad, sugere que existe un efecto del sincretismo de personas en el paradigma verbal sobre la frecuencia relativa de sujetos nulos y sobre la aceptabilidad de oraciones con sujetos nulos en la primera persona del singular: existe una tendencia a evitar la ambigüedad de sujetos nulos con formas verbales ambiguas, pero solo en contextos en los que hay competición de antecedentes. Se analizan estos resultados a la luz de una teoría general de la resolución de la anáfora (ARIEL, 1990, etc.), como resultado de un cálculo que tiene en cuenta la accesibilidad de los antecedentes potenciales y el costo de la forma anafórica (ALMOR, 1996).

PALABRAS CLAVE: Sujeto nulo. Paradigma verbal. Ambigüedad. Resolución de la anáfora.

1 INTRODUCTION

The present paper is concerned with the inflectional system and the use of null and pronominal first-person singular subjects in Brazilian Portuguese [henceforth, BP]. In some languages, agreement-marked verbs may be redundant with the expression of their respective pronominal subjects. However, in many languages, verbal paradigms present ambiguous (syncretic) forms, which do not explicitly reveal the subject of the sentence. Example (1) illustrates this situation in BP. The first verb in the sentence, *dei* “give.PST.1SG”, explicitly marks the subject in the inflection; while the second verb, *estava* “be.IMP.1SG/3SG”, is an ambiguous form, which is compatible with both first and third-person singular subjects.

- (1) como ex-funcionário do Tribunal de Contas, na época inclusive que _ *dei*
 as ex-employee of. the court of accounts at.the time including that _ give.PST.1SG
 a primeira entrevista, eu *estava* no Tribunal de Contas do Estado do Rio de Janeiro.
 the first interview I be.IMP.1SG/3SG at.the court of accounts of.the state of.the Rio de Janeiro.

“As a former employee of the Court of Auditors, at the time at which [I₁] gave the first interview, I₁ was at the Court of Auditors of the State of Rio de Janeiro”(NURC-RJ, “Inquiry_r_52”)

This paper focuses on the interaction of overt morphological verbal markings in BP and their effect on the realization and interpretation of overt and null subjects. The specific goals of this paper are two-fold: (i) to investigate the relation between the verbal paradigm and the use of null and overt first-person singular subjects using quantitative methods; and (ii) to offer a theoretical account for the findings, which attempts to integrate two competing approaches to explain preferences of choice between overt pronouns and null subjects.¹ The first is the Generative account, in which most of the literature on BP null subjects has been couched for the past 40 years (see, for instance, DUARTE, 1995; KATO; NEGRÃO, 2000; *inter alia*). The second is the anaphora-resolution account, which is widely used in the study of null and pronominal subjects in other languages (see GIVÓN, 1983; ARIEL, 1990; ALMOR, 1996; DE LA FUENTE *et al.*, 2016, *inter alia*).

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*).² When BP is compared to other standard pro-drop Romance Languages (for instance, European Portuguese [EP], Spanish and Italian), it appears that the linguistic contexts in which null subjects are allowed are indeed scarcer (DUARTE, 1995; KATO, 1999; BARBOSA *et al.*, 2005, *inter alia*). These restrictions on the use of null subjects are taken to be a by-product of the impoverishment of morphological markings in BP, (often referred to as “Taraldsen’s generalization”, TARALDSEN, 1980; RIZZI, 1986; ROBERTS, 2014; SIMONENKO *et al.*, 2017).

In the present paper, the first-person singular ambiguous (syncretic) and exclusive (non-syncretic) forms are used to investigate the effect of verbal morphology on the use of null and overt subjects. The data discussed here are taken to be a synchronic cut in the history of BP. The results found in both the corpus and the experimental studies do not seem to fully support previous theories on BP. An alternative approach, based on the theory of Anaphora Resolution, is presented and shown to better explain the data.

2 AGREEMENT INFLECTION IN PRESENT-DAY BRAZILIAN PORTUGUESE

BP has substantially modified its inflectional system when compared to previous stages of the language, to other varieties of Portuguese or to other Romance Languages (Italian and Spanish, for instance) (see DUARTE, 1995; KATO, 1999; KATO; NEGRÃO 2000, *inter alia*). Table 1 below summarizes the paradigm of verbal inflectional markings in present-day BP. In Table 2, the verbal paradigm of earlier stages of BP is presented for the sake of comparison (these forms were used at the beginning of the XIXth century).³ In both tables, the forms are given for the verb *falar*, ‘to speak’.

¹ See, e.g., Carminatti (2002) for a similar attempt to combine these two traditions.

² 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.

³ See the Tycho Brahe project (available at <http://www.tycho.iel.unicamp.br/tycho/prfpml/fase2/index.html>) for an overview of the multiple changes in BP during the past two centuries.

Person and Number	TT (exclusive)	TT (ambiguous)
1st singular	<i>falo</i>	<i>falava</i>
2nd singular	<i>fala</i>	<i>falava</i>
3rd singular	<i>fala</i>	<i>falava</i>
1st plural (a gente)	<i>fala</i>	<i>falava</i>
1st plural (nós)	<i>falamos</i>	<i>falávamos</i>
2nd plural	<i>falam</i>	<i>falavam</i>
3rd plural	<i>falam</i>	<i>falavam</i>

Table 1: The Verbal Paradigm in Present-day BP

Source: adapted from Soares (2017, p. 20)

Person and Number	TT (exclusive)	TT (ambiguous)
1st singular	<i>falo</i>	<i>falava</i>
2nd singular	<i>falas</i>	<i>falavas</i>
3rd singular	<i>fala</i>	<i>falava</i>
1st plural	<i>falamos</i>	<i>falávamos</i>
2nd plural	<i>falais</i>	<i>falavais</i>
3rd plural	<i>falam</i>	<i>falavam</i>

Table 2: The Verbal Paradigm in Earlier Stages of BP

Source: adapted from Soares (2017, p. 20)

As shown in Table 1, present-day colloquial BP presents three forms for the present tense and two for the imperfect past tense if, for the first-person plural, only the form *a gente fala/falava* (literally “the people speak/spoke”) is taken into consideration (the *nós* forms, with the *-mos* suffix, are infrequently used in present-day colloquial BP; they are typical of higher register). The main difference between this paradigm and earlier stages of BP, written prescriptive BP, and EP is that second-person singular and plural markings (*-s* for singular and *-is* for plural), as well as first-person plural (*-mos*), are standard in those varieties.⁴ The impoverishment in verbal morphology of present-day BP is well reported in the literature (see DUARTE, 1995; KATO, 1999; KATO; NEGRÃO, 2000, among others). However, it is rarely pointed out that there is variation in the first-person singular according to the verbal tense. Specifically, there is a unique, distinctive inflectional marking of first-person singular in the present (*falo*) and preterite (*falei*) tenses and in those ‘compound’ tenses which are built using the present and preterite tenses of the auxiliary verbs “ter” and “ir” (*tenho falado* and *vou falar*, for example); all other tenses are not exclusively marked for first-person singular, as in, for example, the imperfect (*falava*), the conditional (*falaria*), the periphrastic conditional (*ia falar*), in the indicative; and in the present (*fale*), the future (*falar*) and the imperfect (*falasse*), in the subjunctive; as well as in all the remaining compound forms with the auxiliary *ter* “have” (*tinha falado*, etc). In this light, we distinguish two Tense Types: those in which there is a distinctive inflectional form between first and third person will be called ‘exclusive’ (i.e., non-syncretic); those in which there is no formal distinction will be called ‘ambiguous’ (i.e., syncretic).

Null subjects in current spoken BP combine with any of the verb forms in Table 1 above, although they are becoming rare in certain discourse persons. Kato (1999), for instance, has pointed out that the first and second persons and the referential third are almost never null. Many other researchers have reported, however, that the change in BP has not equally affected each discourse person. This was first pointed out by Negrão (1990), a study based on an oral corpus collected in a public school in São Paulo. Similarly, based on a study of popular written plays, Duarte (1993, 1995, 2000), shows an asymmetry across discourse persons over the period in which BP has become different from EP. Duarte’s hypothesis is that the impoverishment of the inflectional paradigm, shown in Table 1, along with the deactivation of the “Avoid Pronoun Principle” [APP], caused an increase in the number of overt subjects from 20% in the second quarter of the XIXth century to 74% in the 90s.⁵ However, she points out that the increase in the number of overt pronominal subjects is far more drastic in the first and second person than in the third person, as summarized in Figure 1 below.

⁴ The forms in earlier stages of BP shown in Table 2 mostly overlap with EP and prescriptive written BP. It is sufficient to know that, despite a few exceptions (due to syncretic forms), there exists one exclusive marking for each person and number.

⁵ The Avoid Pronoun Principle is a principle according to which pronouns should not be used whenever they are not required (Chomsky 1981).

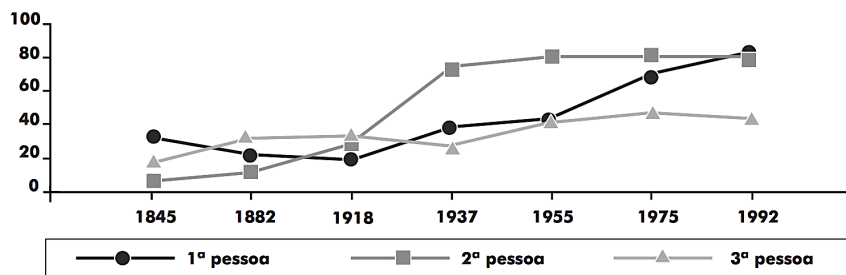


Figure 1: Percentage of Overt Subjects according to Discourse Person

Source: Duarte (2015, p. 97)

As Jeffrey Runner (p.c.) has pointed out, it is clear that the use of overt subjects with the second person increased precipitously after 1918. This coincides with the decrease (and virtual disappearance) of specific inflectional morphological marking in verbs in the second discourse persons. However, the increase in the relative number of first-person overt subjects, which clearly occurred from 1955 on, cannot be linked to impoverishment, since the first-person singular has maintained the same inflection patterns, as shown in Tables 1 and 2.

Duarte (1995, p. 48) explicitly notes that the impoverishment in the inflectional paradigm does not produce gradual and uniform effects across each discourse person. She suggests that the imbalance observed may be a trace of the previous status of BP as a pro-drop language resisting the deactivation of the APP, but she does not propose a more specific explanation. Any attempt to base such an explanation on redundancy (i.e., the idea that there is a preference for avoiding redundancy, so that subject pronouns are avoided as long as they are redundant with inflectional markings, see, e.g., Fernandez-Soriano, 1989) runs into problems. It cannot account for the change affecting the first person, as there is no corresponding decrease in redundancy of inflectional marking since the inflectional marking has not changed over time. The purpose of this paper is to see to what extent further corpus research and psycholinguistic experimentation might shed some light on the question.

3 NULL SUBJECTS AND IMBALANCED DISTRIBUTION IN GENERATIVE GRAMMAR

The idea of the so-called pro-drop parameter was originally introduced into the Generative tradition by Perlmutter (1971), based on null subject languages like Spanish, Italian, Serbo-Croatian, etc. The distinction between pro-drop and non-pro-drop languages was, at least for some languages and some approaches, claimed to be a difference as to how the subject position is filled: a pronoun in subject position, which is either deleted or unpronounced in null-subject languages, or obligatorily pronounced in non-pro-drop languages (see, e.g., CHOMSKY 1981). The generalization about the impoverishment of the agreement paradigm and the absence of null subjects initially appeared robust across many languages (see ROBERTS; HOLMBERG, 2010; ROBERTS, 2016; for overviews).

As early as the 80s, the correlation between rich agreement morphology and the pro-drop parameter was called into question (see GILLIGAN 1987, for an overview of this discussion in the 80s). Basically, the existence of languages in which co-referential null subjects are permitted with verbs whose agreement morphology is completely absent, such as Japanese, Mandarin and Thai, challenges the parametric co-occurrence of rich verbal morphology and null co-referential subjects. The concept of 'impoverishment' has been reformulated a number of times in terms of richness of agreement morphology (RIZZI, 1982), morphological uniformity (JAEGGLI; SAFIR, 1989), undistinctive morphology (ROHRBACHER, 1994) and syncretism (RIZZI, 2002). However, mixed systems, such as that of BP, always create a problem for such approaches and certainly counting and comparing the number of ambiguous forms across languages, as proposed in some of them, does not provide a reasonable account for the observed data.

BP has played a major role in this discussion since the 1980s, when Tarallo (1983) claimed that it was undergoing a change in the pro-drop parameter. During the 1980s and 1990s, much research further contributed to this claim, suggesting that BP is a “live sample” of parametric change or a “partial pro-drop language”.

In general, the crucial motivation for the change of BP toward non-pro-drop status has been claimed to be the impoverishment of verbal inflectional morphology, as described in Section 2. Recent literature proposes two possible explanations as to how this impoverishment has affected the possibility of having a null subject in BP: (i) because it is no longer distinctive and strong, the inflectional marking does not have the proper features to license the null pronoun in subject position (HOLMBERG *et al.*, 2009); (ii) differently from standard pro-drop languages, in which the inflectional marking is the subject and satisfies the syntactic requirements of the sentence, the inflectional marking in BP is no longer sufficient to check the features of subject position, such as the EPP feature, and an overt pronoun is obligatory to fill the position and check the relevant features (KATO, 1999; BARBOSA *et al.*, 2005).⁶ Assuming either position suppose that these requirements apply across the board. That is, one would not expect that any specific tense or any person is more likely to be expressed with overt pronouns than any. Among the various positions in the literature trying to explain the change, none really allow to understand why there is an imbalance across persons.

4 ANAPHORA RESOLUTION AND NULL SUBJECTS

In the literature on anaphora resolution, it is generally assumed that there is a reverse mapping principle between antecedent accessibility (or salience) and anaphor explicitness that guides resolution: more accessible (salient) antecedents are retrieved by less complex and less informative anaphoric forms (ARIEL, 1990, 1994, 2001; GUNDEL *et al.*, 1993; GROSZ; SNIDER, 1986; ALMOR, 1996, 1999; among many others).⁷ Earlier functional approaches observed that in English, for example, an unstressed personal pronoun is more likely to refer to the referent of the object of the preceding sentence than to the referent of a complement within that object, for which a full NP is preferred, as shown in (3).

- (2) I bought a computer with an external battery,
 (a) but it/#the computer was stolen. [it = the computer]
 (b) but the battery/#it was stolen. [it = the battery]

In her particularly influential 1990 book, Ariel proposes a detailed relationship between the complexity/explicitness of the referential form and the accessibility of the antecedent, which she calls the Accessibility Marking Scale (ARIEL, 1990, p. 73). The scale was further elaborated by Arnold (1998, p. 19), whose version is given in Figure 2.

As shown in Figure 2, the Scale of Accessibility provides an account for the preferences noted in (3): the intuitively most accessible referent in (3), the computer, is preferably retrieved by an unstressed, lower ranked pronoun in continuation (3a); in continuation (3b), a less accessible antecedent, the battery, is preferably recovered by a short definite description, which is higher in the hierarchy as it is more explicit and complex.

For the BP facts under investigation here, the four lowest levels are relevant. Basically, Ariel (1990) states that null pronouns are less complex and explicit than overt pronouns so that they have a very strong bias toward the most accessible antecedents.

⁶ The EPP feature is the Minimalist Program version of Chomsky (1981)'s Extended Projection Principle, according to which every sentence must have a subject.

⁷ For the purposes of the present paper, we do not need to distinguish between the various definitions of accessibility and salience that have been proposed in the literature mentioned above, as they all share the common relevant idea that less specified forms are used to refer to more accessible / more salient antecedents.

	Marking Scale	Examples	
L o w	Full name + modifier	Joan Smith, the president	
	Full ('namy') name	Joan Smith	
	Long definite description	The tall and authoritative president	
	Short definite description	The president	
	Last name	Smith	
	First name	Joan	
	H i g h	Distal demonstrative + modifier	that hat we bought last year
		Proximal demonstrative + modifier	this hat we bought last year
		Distal demonstrative + NP	that hat
		Proximate demonstrative +NP	this hat
Distal demonstrative		that	
Proximate demonstrative		this	
a c c e s s i b l i t y		Stressed pronoun + gesture	SHE (plus gesture)
		Stressed pronoun	SHE
		Unstressed pronoun	she
		Cliticized pronoun	(no examples in English)
	Extremely High Accessibility Markers	gaps, including pro, PRO and <i>wh</i> -traces, reflexives, and Agreement	

Figure 2: Ariel (1990)'s Scale of Accessibility as revised by Arnold (1998)

Source: Arnold (1998, p. 19)

5 REASSESSING THE CORPUS DATA

In this section, the results of a reanalysis of a corpus of oral interviews (NURC-RJ) are reported. This corpus was previously analyzed by Duarte (1995). The new analysis was carried out for two main reasons: (i) from a data perspective, some of the criteria used to exclude certain data were too restrictive, such as the fact that no “contrastive” subjects were taken into consideration; and (ii), from a technical perspective, thanks to new analytical toolkits, such as new statistical packages, relevant factors and correlations might be discovered that could not have been discovered at the time (see, for instance, GRIES, 2015, for a critical point of view on previous corpus studies without inferential statistical analysis and for arguments in favor of using (generalized) linear mixed models in this sort of analysis). Nine interviews carried out in the 70s and nine interviews from the 90s were analyzed. Overall 8032 inflected clauses were gathered. We included the cases in which the subject was either co-referential or generic (following DUARTE, 1995, cf. p. 36) but also the cases of categorical null subjects (e.g., affirmative answers and fixed expressions) and categorical overt subjects (e.g., contrastive subjects, pronouns with numerals, such as *nos duas* ‘the two of us’, etc.), which Duarte excluded. Null expletive and presentational third-person subjects were excluded, in line with Duarte.

5.1 METHODOLOGICAL CONSIDERATIONS

The corpus research presented here was carried out using the following procedure in order to guarantee unbiased results. First, eighteen texts were selected from the NURC-RJ corpus. The criterion used to select these texts was the availability of the audio file, so that in dubious cases the primary source of the data would be available for verification (about 0.5% of the data was in fact verified). The second step was automatically annotating the transcriptions (also available on the NURC-RJ website). The automatic parser

VISL-Portuguese was used to fully annotate the eighteen texts.⁸ The sentences with null subjects were manually collected⁹. Afterwards, the sentences with full pronominal forms were collected through searches on each form (*eu, tu/você*, etc). These data were descriptively analyzed in qualitative and quantitative terms. Finally, an inferential analysis was carried out using logistic regressions with the *glmer* function of the LanguageR package in R, applying the logit linking function with Laplace approximations¹⁰.

The analysis started by splitting the cases according to the standard classification by discourse person of the subject. This was followed by a classification of the verbal inflection according to its Tense Type, taking into consideration only the agreement markers which are explicitly marked in present-day BP (cf. Table1 above)¹¹.

5.2 RESULTS

The results found in the present study are generally compatible with those found in the previous literature (e.g., NEGRÃO, 1990; DUARTE, 1995). 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 discussed for Figure 1 above, first-person subjects differ according to Tense Type, as shown in Table 1. 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, as shown in Figure 3 below.

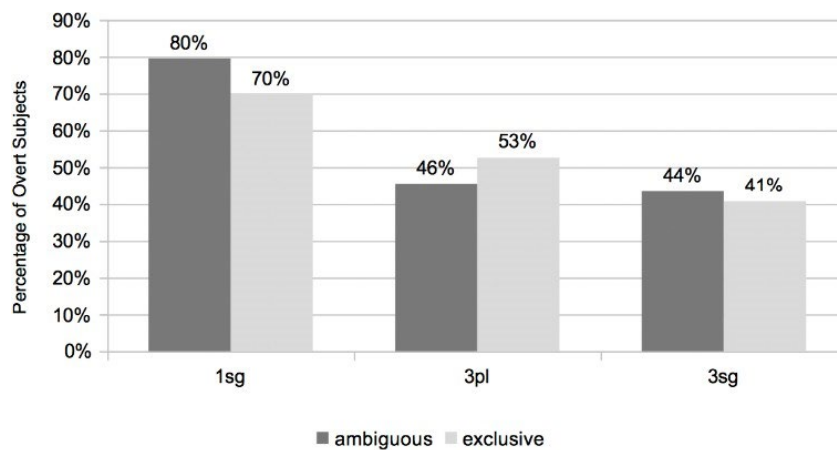


Figure 3: Percentage Overt Subjects according to Overt Inflectional Marking

Source: Soares (2017, p. 50)

⁸ More precise information about the annotation tool is available on the website of the project: <http://visl.sdu.dk/visl/about/>. This tool is part of the project *Floresta Sintá(c)tica* ("Syntactic Treebank"), a collaboration between Linguateca and the VISL project. It contains texts in Portuguese (from Brazil and Portugal) annotated (analyzed) automatically by the parser PALAVRAS (BICK 2000) and reviewed by linguists.

⁹ The first author of this paper collected the clauses with null subjects. Each clause was further sorted according to twelve independent criteria (only two of which are reported in this paper). This means that each case was examined twelve times (guaranteeing the quality of the data).

¹⁰ The models included the interviewed speaker as a random factor, as suggested by Gries (2015). See also Baayen *et al.* (2008), Bates and Maechler (2009), Bates *et al.* (2015) for details.

¹¹ Due to an oversight, incidentally lexically ambiguous past verb forms from verb classes where they are not usually ambiguous (such as *soube* "know.PST.1SG/3SG", *disse* "say.PST.1SG/3SG", *quis* "want.PST.1SG/3SG" and *trouxe* "take.PST.1SG/3SG"), were not classified in the ambiguous verb tense group as they should have been. We did not take the trouble to correct this as such occurrences add up to only a very small number of cases (less than .1% of the data).

In order to correctly interpret Figure 3, it should be noted that 1sg, 3pl and 3sg denote discourse persons (not inflectional markings). Ambiguous vs. exclusive marking distinguishes those Tense Types in which the inflectional marking is the same (ambiguous) or different (exclusive) for the first singular and third singular discourse persons. This same distinction is applied for all three discourse persons reported here although exclusive marking only concerns 1sg. It is important to keep in mind, however, that 3sg person marking is always ambiguous between at least the second and third discourse persons. Also, for 3pl, there is no possible ambiguity with the 1st person. Exclusive vs. ambiguous refers to the distinction between Tense Types that is relevant to the 1sg and 3sg.

The number of first-person singular overt subjects is overall higher than the third-person singular and plural. Ambiguous (syncretic) marking seems to have a small influence on the overall number of overt pronouns: across condition, about 5% more subjects were overt in ambiguous verbal Tense Types than in exclusive verbal Tense Types. Noticeably, the effect of the ambiguous Tense Types is not uniform across inflectional markings:

First-person singular subjects were more frequently overt when followed by a verb in an ambiguously marked Tense Type than when followed by a verb in a non-ambiguously marked Tense Type. This 10% difference suggests that the ambiguously marked Tense Types favors the use of overt subjects in the first-person singular, although it does not transparently affect the third persons in the same way. For third person, the effect is inverted in 3pl and is numerically weaker for 3sg.

A logistic regression model with Discourse Person (1sg, 3sg, 3pl) and Verbal Tense Type (ambiguous, exclusive) as fixed Factors indicates that the person marking contributes in a significant way to the choice between overt and null subjects (first-person singular vs. third-person plural: β : -2.055 SE: 0.30 z-value: -6.850 and p-value: 7.38e-12; first person singular vs. third-person singular: β : -2.265 SE: 0.186 z-value: -12.180 p-value: < 2e-16). We also found a significant main effect of the factor Verb Tense Type (β : -0.648 SE: 0.149 z-value: -4.349 p-value: 1.37e-05) with fewer overt pronouns for exclusively marked tenses. A significant interaction was established between Verb Tense Type and Discourse Person, presumably because Verb Tense Type showed an inverse pattern for third-person plural as regards the Tense Type (more overt subjects in exclusively marked verbal tenses) compared to first and third-person singular (β : 0.613 SE: 0.195 z-score: 3.146 and p-value: 0.00165). Finally, the intercept term, which was the first-person singular in ambiguously marked tenses, was significant (β : 2.662 SD: 0.216 z-value: 12.327 p-value: < 2e-16), showing a preference for the use of overt subjects.

5.4 DISCUSSION

An effect of inflectional marking was found: when the first-person singular is used in a Tense Type that does not exclusively reveal the discourse person of the subject, the overt form is preferred to the null form. The combination of first-person singular together with an ambiguous Tense Type is statistically significant. Importantly, overt subjects were more frequent for first-person singular in ambiguous Tense Types (as predicted by an impoverished morphology hypothesis) but also were they for exclusive Tense Types. In this, the corpus findings suggest that impoverishment has led to a bias in favor of maximal informativeness in present-day BP. The effect of Verb Tense Type on third persons did not show a clearly interpretable pattern with respect to the impoverishment hypothesis¹².

6 EXPERIMENTAL EVIDENCE

Given the results obtained in the corpus research, two experiments were carried out in order to test whether ambiguous markings influence the acceptability of null and overt subjects in a more controlled environment. In Experiment 1, exclusive vs. ambiguous verbal marking was tested by varying the tense of a subordinate clause verb whose subject was always first-person singular. The results of the first experiment suggested that one aspect of the design, the presence of competing animate antecedents, might have

¹² The apparently contradictory effect for third persons can in fact be explained away in a broader context, though it is beyond the scope of this paper to do so. See Soares (2017, chap. 1).

been relevant. This led us to run a second experiment where materials were designed to rule out any effects of competition. The results of the experiments partially confirm the predictions made from the corpus analysis but suggest that the crucial factor is ambiguity of reference, rather than ambiguity of marking *per se*.

6.1 EXPERIMENT 1: EXCLUSIVE VS AMBIGUOUS MARKING

The purpose of Experiment 1 was to adjudicate between the predictions made by classical generative accounts and the findings of the corpus investigation, by checking whether the choice between ambiguous and exclusive Tense Types does or does not have an effect on the acceptability of null and overt subjects. Specifically, the corpus study would lead us to expect an interaction, with null subjects being more acceptable with exclusive Tense Type and overt subjects being more acceptable ambiguous Tense Type. By contrast, the classical generative accounts lead us to expect the acceptability of overt subjects to be higher across the board, irrespective of Tense Type, as redundancy is not assumed to continue to have an effect when the system becomes impoverished.

6.1.1 Material design

The purpose of this experiment was to investigate whether the preference for a null or an overt 1st person singular subject pronoun is influenced by inflectional marking on the verb, specifically by whether the Tense Type is ambiguous or exclusive. The materials were thus constructed on the basis of two binary factors, namely the status of the subject, which can be null or overt, and the status of the Tense Type, which can involve exclusive or ambiguous marking:

- (i) Tense Type: exclusive vs. ambiguous
- (ii) Subject: null or overt

This led to the four conditions illustrated in the following segments:

- (a) exclusive+null: Eu divulguei ('I published.1SgPst')
- (b) exclusive+overt: Divulguei ('published.1SgPst')
- (c) ambiguous+null: Eu ia divulgar ('I was.going.1/3SgImpft to.publish')
- (d) ambiguous+overt: Ia divulgar ('was.going.1/3SgImpft to.publish')

Each experimental item took the form a short dialogue. Speaker A's turn provides a context making the relevant experimental sentence, uttered by Speaker B, sound as natural as possible. This context was composed of two sentences: an introductory sentence followed by a second sentence with an indirect temporal interrogative introduced by *quando* 'when', asking when an event took place. Speaker B's answer to that question displayed the four conditions in a temporal adjunct subordinate clause, also introduced by *quando*, as in (3) below:

(3) **Context:**

A – A Maria estava muito nervosa. Você sabe quando ela ficou mais calma?

"Mary was very nervous. Do you know when she's got calmer?"

Conditions (a) and (b)

B – Eu_i tranquilizei a Maria quando eu_{i/_1} divulguei os resultados do exame.
I calm.down.PST.ISG the Maria when I publish.PST.ISG the results of.the exam.

"I calmed Mary down when I published the results of the exam."

Conditions (c) and (d)

B – Eu_i tranquilizei a Maria quando eu_{i/_1} ia divulgar os resultados do exame.

I calm.down.PST.ISG the Maria when I was.going.to publish.INF the results of.the exam.

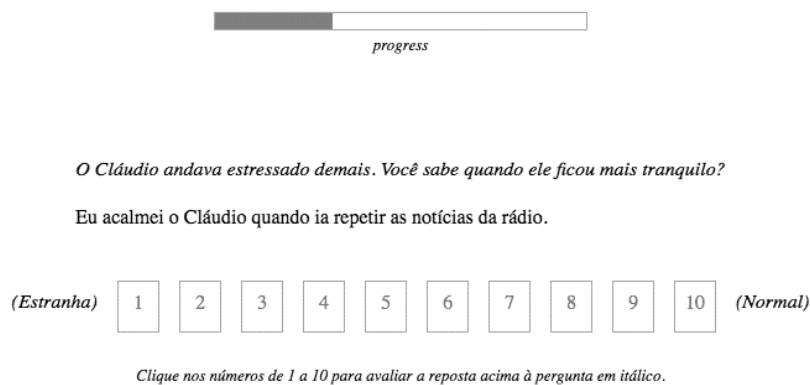
"I calmed Mary down when I was going to publish the results of the exam."

Twenty-four items of this type were created for this experiment, based on a previous experiment carried out for different purposes (FERNANDES *et al.*, 2018)¹³.

6.1.2 Procedure

Participants read the sequence of two turns. They were asked to judge the acceptability of the answer in the context provided on a Likert scale from 1 to 10, cf. Figure 4 for a typical stimulus.¹⁴ They were told to use the full scale according to how natural (“Normal”) or strange (“Estranha”) the answer seemed in the context of the question. After judging acceptability, the participants were asked about the interpretation of the relevant subject – null or overt – in a closed yes-no question task, cf. Figure 5.

Figure 4: Screen sample – Judgment Task



progress

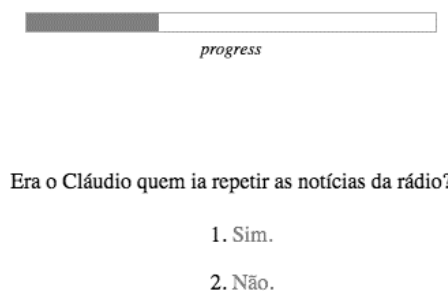
O Cláudio andava estressado demais. Você sabe quando ele ficou mais tranquilo?

Eu acalmei o Cláudio quando ia repetir as notícias da rádio.

(Estranha) (Normal)

Clique nos números de 1 a 10 para avaliar a resposta acima à pergunta em itálico.

Figure 5: Screen Sample – Closed Question Task



progress

Era o Cláudio quem ia repetir as notícias da rádio?

1. Sim.

2. Não.

All participants voluntarily participated in the experiments on the IbeXFarm platform ([http:// spellout.net/ibexfarm](http://spellout.net/ibexfarm), DRUMMOND, 2014). The experiment started with four practice items and took about 30 minutes to complete. The target-items were randomly interspersed with 52 distractors, in such a way as to insure that any two target-items were separated by at least two

¹³ Fernandes *et al.* (2018) showed that the specific context provided for these items is heavily biased with respect to the preferred interpretation of the empty subject of the subordinate clause in B's answer: on average, 80% of responses to an antecedent choice task were for the subject of the matrix, *viz.* the first-person singular in the present case. The verbs used in the subordinate clauses were explicitly chosen to not show any intrinsic implicit causality bias: outside the specific context of this experiment, all of them were ambiguous as regards their bias towards the agent or patient of the preceding clause. Other item-related properties were addressed statistically in the models reported below, taking items as a random factor.

¹⁴ A reviewer for FL asked why we chose a Likert scale ranging from 1 to 10. The scale was chosen to facilitate the task for Brazilian participants, who are accustomed to having school work graded on that scale and typically report finding it easier to use. In a cross-cultural study, Lee *et al.* 2002 show that the choice of range for Likert scales has no effect on the results of questionnaires (patterns across conditions are maintained regardless of the range of the scale), but that choosing a range that is culturally well-established can significantly decrease the difficulty of the task for participants (evidenced by a decrease in the number of skipped items and out-of-range responses).

distractors. Each participant was presented with a single condition of each experimental item using a Latin-square design crossing participants and items. Among the items, four perfectly acceptable control sentences were inserted. Four control sentences that violate strong grammatical or pragmatic constraints were inserted at the end of the experiment in order to insure that participants were attentive until the end and to avoid ceiling effects for the experimental items.

6.1.3 Participants

Twenty-seven participants took part in this experiment, all of them highly educated (minimally under-graduate students) living in the southern region of Brazil (Rio Grande do Sul and Santa Catarina). They were invited to participate via email and social networks. Before starting the experiment, the paradigm was explained to the participants and they filled in a basic information form that included a declaration of informed consent. Their age averaged 38,5 years (± 7 years). In the analysis, the results of participants who either scored below 80% in the interpretation task or rated ungrammatical control sentences above eight were discarded (three participants). Only acceptability judgments for which a correct answer was given to the control question were taken into account (leading to the exclusion of four data points).

6.1.4 Empirical predictions

The empirical predictions are: (i) if the participants principally rely on verbal marking, an ambiguous Tense Type along with a null subject should be judged worse than the other three conditions; (ii) if the participants take overt subjects to be redundant with verbal exclusive marking, the overt subject should be judged worse than the other three conditions for being redundant (over-informative) with the exclusive marking; (iii) if participants principally care to minimize redundancy while avoiding ambiguity, we should expect an interaction with overt subjects being judged more acceptable with ambiguous Tense Type and null subjects judged more acceptable with exclusive Tense Type; and (iv) if the participants take the system as a whole to be impoverished and thus structurally and informatively deficient, overt subjects must be preferred across the board, but no interaction with Factor TT would be expected.

6.1.5 Results

As shown in Figure 6 below, the combination of an ambiguous Tense Type with a null Subject reduces the acceptability of the sentence (averaging 6.7) when compared to all the other conditions (ambiguous Tense Type with overt subject, mean 7.5; exclusive Tense Type with overt subject, mean 7.6; and exclusive Tense Type with null subject, mean 7.7)¹⁵

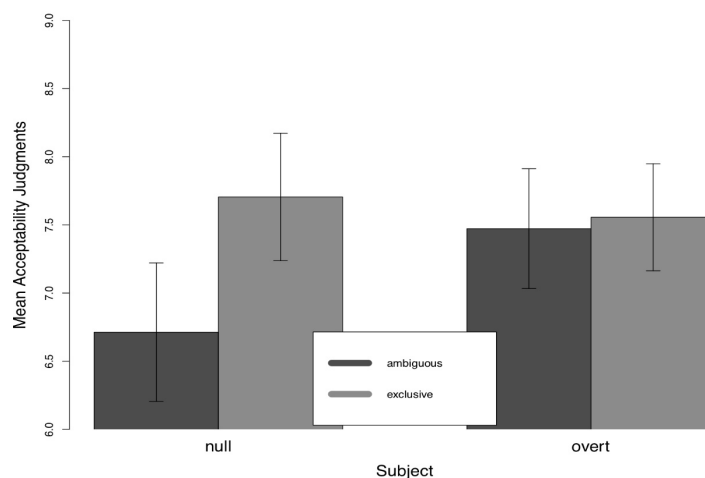


Figure 6: Experiment 1 – Mean Acceptability Judgments according to Tense Type and Subject Factors

Source: adapted from Soares (2017, p. 64)

¹⁵ In Figure 6 and all the following graphs, error bars represent 95% confidence intervals.

For the inferential statistical analysis, raw acceptability judgments were centered and re-scaled in a by-subject analysis (z-score transformation). They were entered into a log-linear mixed-effects model analysis containing two Factors (Tense Type and Subject) with two levels and random effects (Participants and Items), including random slopes (BARR *et al.*, 2013). The outcome of the model using z-scores instead of the raw data, which showed the best distribution of residuals, is summarized in Table 3 below.

Factor	Estimate	Std. Error	T-value	Pr(> t)
(Intercept)	-0.3176	0.1181	-2.690	0.01245 *
Subject	0.3435	0.1459	2.354	0.02699 *
TT	0.4692	0.1422	3.300	0.00294 **
Subject:TT	-0.3641	0.1782	-2.043	0.04548 *

Table 3: Log-linear mixed-effects model for Judgments in Experiment 1 (z-score transformation)

The model shows a significant effect of the interaction and of both main Factors. These results show that null subjects with ambiguous TTs are significantly less acceptable than other conditions.

6.1.6 Discussion

The results of Experiment 1 confirm the results of the corpus analysis in one sense: BP speakers tend to dislike the ambiguity generated by an ambiguous verbal marking preceded by a null subject. This claim is supported by the higher number of overt subjects found in the corpus in the 1st person singular in the ambiguous Tense Type and by the drop in acceptability in Experiment 1 when the verb was neither exclusively marked nor preceded by an overt subject. It appears that BP speakers have no dispreference for redundant marking but that ambiguity is clearly dispreferred. These results thus suggest that BP speakers have paradigmatic knowledge that the ambiguously marked construction with a null subject is not optimal for comprehension and that the most informative and efficient form, in this case, is the overt subject. However, the particularly low acceptability of null subjects with ambiguous verb forms may have another reason. All our target items had two animate referents, which were potential antecedents of the null subject. Although participants interpreted the items as intended in the items we included in our data set, this ambiguity may have resulted in reduced acceptability. We ran a second experiment to shed light on this possibility.

6.2 Experiment 2 — The relevance of Verbal Marking without competing antecedents

Taking into consideration the results obtained in Experiment 1 and the observations from the corpus, Experiment 2 is a follow-up study with slightly different materials. Instead of having two potential antecedents for the null and overt subjects, items were set up with inanimate direct objects in the main clause, so that only the subject of the main clause could be reasonably interpreted as the antecedent of the subject in the temporal subordinate clause. With this set-up, it is expected that no effect of competition could show up so that only the effect of the verbal paradigm and the use of null or overt subjects is at stake. The same two experimental Factors were tested. The hypotheses about the interaction between Tense Type and null and overt subjects were the same.

6.2.1 Materials

Twenty-four items were created based on those of Experiment 1. The same context sentences were provided. After this context sentence, a temporal question introduces the answers in (3), which display the four experimental conditions. As in Experiment 1, the verb was either preceded by a null or overt subject (Factor Subject) and the verbs were in a tense that either was exclusive for the first-person singular or ambiguous between first and third-person singular (Factor Tense Type). The only difference between Experiment 1 and 2 is that in the latter direct objects were always inanimate, and thus were implausible candidates as antecedents for the subjects in the temporal clause. After judging the short dialogue composed of these two turns, participants answered an interpretation task, identical to that in Experiment 1.

(3) **Context:**

A – A Maria estava muito nervosa. Você sabe quando ela ficou mais calma?

“Mary was very nervous. Do you know when she’s got calmer?”

Conditions (a) and (b)

B – Eu₁ resolvi o problema quando eu₁/1 divulguei os resultados do exame.

I solve.PST.1SG the problem when I publish.PST.1SG the results of.the exam.

“I solved the problem when I published the results of the exam.”

Conditions (c) and (d)

B – Eu₁ resolvi o problema quando eu₁/1 ia divulgar os resultados do exame.

I solve.PST.1SG the problem when I was.going. to publish.INF the results of.the exam.

“I solved the problem when I was going to publish the results of the exam.”

6.3.2 Participants and methods

Twenty-four subjects, different from those in Experiment 1, took part in this experiment. They were on average 37.1 years old (± 7 years). In the analysis, two participants were discarded because they scored the ungrammatical control sentences above 8. Otherwise, the procedure was the same as in the previous experiment.

6.3.3. Empirical predictions

These are the same as in Experiment 1, see section 6.1.4 above.

6.3.3 Results

The only relevant significant effect was an overall dispreference for overt pronouns.¹⁶ The interaction between the two factors was not significant. (Numerical means for the conditions were as follows: ambiguous Tense Type and null subject: 7.2; ambiguous Tense Type and overt subject: 6.7; exclusive Tense Type and null subject: 7.6; exclusive Tense Type and overt subject: 7.4.)

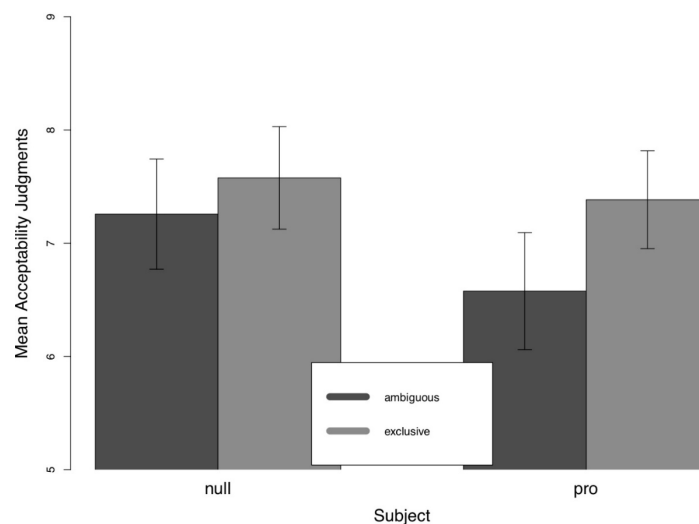


Figure 7: Experiment 2 – Mean Acceptability Judgments according to Tense Type and Subject Factors

¹⁶ There is also an overall dispreference for the ambiguous Tense Type, but this is clearly an orthogonal property of the materials, which were constructed with the past tense in the main clause. Thus, choosing the unambiguous Tense Type in the subordinate clause leads to maintaining the same tense (past) in both clauses, whereas choosing the ambiguous Tense Type leads to switching from past in the main clause to the periphrastic conditional or a compound form with the verb *ter* “have” in the subordinate clause, which is clearly dispreferred.

As for the inferential statistical analysis, as in Experiment 1, the z-score transformed judgments were entered in a linear mixed-effects model. The main Factors were Subject and Tense Type, and the random Factors were Participants and Items, including random slopes. The maximal model is summarized in Table 5 below.

Factor	Estimate	Std. Error	T-value	Pr(> t)
(Intercept)	0.00341	0.09956	0.034	0.9728
Subject	-0.24239	0.10856	-2.233	0.0260 *
TT	0.20679	0.11021	1.876	0.0611 .
Subject:TT	0.06613	0.15436	0.428	0.6685

Table 4: Log-linear mixed-effects model for Judgments in Experiment 2 (z-score transformation)

As can be seen in Table 4, the main Factor Subject is significant and the main Factor TT is marginally significant. Crucially, the interaction between the main Factors does not reach significance.

6.3.4 Discussion

The results of Experiment 2 show a main effect of the Factor Subject, but it is not modulated by the Factor Tense Type: in contexts where there is only one possible antecedent for an anaphoric subject, null pronouns are preferred, regardless of verbal inflection. As was the case for Experiment 1, these results argue against the idea that avoiding redundancy and referential ambiguity is the central explanatory factor, since this would predict an interaction (null would be expected to be more acceptable with exclusive Tense Type and overt would be expected to be more acceptable with ambiguous Tense Type). Note that we find a general preference for null pronouns here, a result that is not predicted at all if Brazilian speakers take the whole system as impoverished. Such impoverishment would have predicted a preference for overt pronouns. What we see here is that in a situation where the subject antecedent is highly accessible, the null pronoun is preferred. It thus seems that accessibility is a better predictor of the choice between overt and null pronouns than any prediction that can be made on the basis of the verbal markings as such.

The results of this experiment also suggest a somewhat different explanation for the effect of ambiguous Tense Type (namely, an increase in overt subjects) found both in the corpus research and in Experiment 1. Specifically, we suggest that it is a by-product of the presence of competing potential antecedents for the anaphoric subject. This might be a topic for future experimental research¹⁷.

7 GENERAL DISCUSSION

Since the 80s, the relation between the impoverishment of verbal morphology and the availability of referential null subjects in a given language has been observed and often assumed to be quite direct (TARALDSEN, 1980; CHOMSKY, 1981; RIZZI 1980, 1982, 1986; ROHRBACHER, 1994; *inter alia*). BP has been considered a key piece of evidence to support such claims (DUARTE, 1993, 1995, 2000, among many others).

In this paper, further corpus and experimental evidence that challenges the correlation between the impoverishment of verbal morphology and the use of null subjects in BP has been presented. An effect of verbal inflection is observed: when first-person singular verbs are ambiguous (syncretic), there is a decrease in the relative number of null subjects in corpora and a parallel decrease in the acceptability of the sentences without overt subjects in Experiment 1. However, as shown by Experiment 2, this effect seems to be a by-product of the competition between potential antecedents for the null subject in contexts where the verb is ambiguous between first and third persons.

¹⁷ Though it might be tempting to consider analyzing the corpus data in terms of the competition between potential antecedents, this turns out to be a methodologically highly complex problem, since it requires access to the background assumptions of the speakers, which are not available. Experiments on the other hand, allow full control of these factors.

In the Generative frameworks assumed by previous literature, these results might be approached in two different ways: (i) the first-person singular exclusive agreement marker is a trace of the previous organization of the BP system and, for this reason, the higher number and the higher acceptability of null subjects with these verbal markings are also a trace of the previous system (DUARTE, 1993, 1995, 2000); or (ii) in the current system of BP, not all null subjects are the same (NEGRÃO; MÜLLER, 1996; FIGUEIREDO-SILVA, 2000; HOLMBERG *et al.*, 2009), and specific verbal inflections can license the null subject (BARBOSA, 2011; NUNES, 2015).

This paper does not aim to analyze the research based on assumption (i) in detail. It seems that Duarte (1995) is right about the deactivation of the APP in BP (to the extent that it ever was active). As mentioned before, the results of the experiments suggest that redundancy in the expression of first-person singular subjects is not dispreferred as such by BP speakers.

On the other hand, the authors who assume (ii) usually claim that the distribution of third-person null subjects depends on whether they are analyzed as, e.g., pronominal anaphors (BARBOSA *et al.*, 2005), bound variables (NEGRÃO; MÜLLER, 1996), silent variables (MODESTO, 2000, 2008a,b) or even traces of movement (FERREIRA, 2004), but none of them has considered the effect of the first-person exclusive vs. ambiguous agreement on the choice between overt and null subjects. Of course, this effect does not constitute an obvious counter-example for these theories, yet there is no obvious way to account for it within their systems.

There are two authors who might possibly offer an account based on (ii) for the facts observed in the present paper. The first possible explanation based on (ii) is Barbosa (2011), building on Modesto (2008a,b) and Holmberg *et al.* (2009). Her account is based on the fact that null subjects are minimally specified NPs (much like a bare noun), whose definite reading must be triggered compositionally by checking a D-feature. Mentioning data from Hebrew and Russian (languages that also have exclusive inflectional markings for some tenses but not for others), she suggests that a type-shifting operation takes place when the D-feature is checked. In a brief outline of future work, she cites Ritter (1995), who proposed that verbal agreement in Hebrew has a D feature only in Past and Futures tenses, and Shlonsky (2009), who claims that first and second-person agreement morphemes in Hebrew are incorporated subject clitics while third-person agreement enters an underspecified person slot. Such proposals seem, *mutatis mutandis*, to provide a possible account for the data on different verbal markings but not for the effect of competitors. Moreover, following Barbosa's reasoning, type-shifting in BP, which is assumed to be necessary to get the null subject, is triggered by movement to a topic position rather than by the specific tense-agreement combinations found to be relevant in the corpus study and experiments. Furthermore, in the corpus data, the effect of Tense Type is not limited to topic pronouns. The second possible explanation is the approach proposed by Nunes (2015). He suggests that the basic value for person in BP is the value default (i.e. what is usually thought of as inflectional third person, corresponding the third and second discourse persons), whereas first-person singular is more complex. In the exclusive forms, it has a complete set of Φ -features [P:1,N:SG] in the T[ense] head, while in ambiguous forms, it has an incomplete set of Φ -features [N:SG].¹⁸ One might attempt to correlate this difference with a preference for null or overt subjects with each set of Φ -features. This would make them avoid first-person singular overt subjects with exclusively marked verbs. But, this is not what was found in Experiment 1, in which a preference for overt subjects with non-syncretic forms was found and, especially, this proposal cannot account for the preference for null subjects regardless of Tense Type evidenced in Experiment 2.

Based on Gilligan (1987) and the evident typological discrepancies across languages, Newmeyer (2004, 2005) claims that the parametric correlation between the richness of agreement morphology and the null realization of subjects (among other parametric correlations) cannot be sustained. Roberts and Holmberg (2010) agree with some of the observations made by Newmeyer (2004, 2005) but simply suggest that the data is rather inconclusive. In an attempt to integrate these various divergent patterns into the framework of the parametric theory, Biberauer *et al.* (2010), Roberts and Holmberg (2010) posit a more complex notion of pro-drop, according to which languages are assumed to be distributed into several subgroups. The notion of "partial" pro-drop emerges from this proposal: Finnish, Russian and BP are hypothesized to be "partial" pro-drop languages. However, if this is the defining feature of a "partial" pro-drop language, BP may not be required to fall under this classification, since the morphosyntactic constraints proposed in the literature are better explained in terms of general anaphora resolution constraints on the interpretation

¹⁸ Nunes (2015)'s proposal is in a way more complex because it aims at accounting for hyperraising facts in some specific dialects of Brazilian Portuguese. According to him, at least some speakers (including himself) have a full set of Φ -features for exclusive forms while others do not. Our interpretation of his proposal might offer some insights into the interpretation of our findings in the corpus but not for the experimental data. The facts about hyperraising are far beyond the scope of this paper (see more details in NUNES, 2015).

of anaphoric subjects. These claims fit in well with what is proposed by Newmeyer (2004, 2005, 2006), according to whom the parsing-based theory of Hawkins (2004, 2014) accounts neatly for many of the generalizations that, in the past, had been attributed to differences in parameter settings: “[M]any patterns of interaction, even in grammars, can be explained when viewed from an efficiency and ease of processing perspective” (HAWKINS, 2014, p. 202). It is not the main aim of this paper to discuss the plausibility and the coverage of the theory of parameters. However, it is clear that the data presented here provide challenging evidence.

The data at stake here can be accounted for in terms of a general theory of anaphora resolution. It is necessary, however, to stress two points of this theory, one concerning the levels of complexity of null subjects in BP and the other related to the accessibility of antecedents in the experimental contexts. Let us start with the different levels in Ariel (1990)’s scale of complexity. In her scale, null anaphors and inflectional markings are not at the same level of Complexity: null subjects that are related to agreement markers, such as those of other pro-drop Romance languages, are higher on the scale (more complex and explicit) than actual “zero expressions”, such as those of pro-drop languages that lack agreement markers altogether (Mandarin Chinese, for instance) (ARNOLD, 1998; DE LA FUENTE, 2015; *inter alia*). It is assumed here that BP has a mixed system, which explicitly and exclusively marks some first persons (these are, thus, higher in Ariel 1990’s scale), while the other null subjects are, as in Chinese Mandarin, real “zero expressions”. This explains the effect of exclusive marking found in the present paper in both Experiment 1a and in the corpus data. As for the Accessibility of the antecedents, based on Givón (1976, 1983), Ariel (1990) points out that many factors must be taken into account when calculating their salience. Givón (1983)’s approach establishes a scale according to which the harder it is to identify the referent (because it is surprising or confusing, or involves some kind of textual discontinuity), the larger the quantity of information encoded in the anaphoric form will need to be. It partially predicts the data in Experiments 1 and 2, but further refinements are possibly necessary. It is worth noticing that Ariel (1990, 1994, 2001) incorporates this very same factor (“competition”), along with others, as a predictor in her Accessibility Scale.

Almor (1996, 1999, 2000) also incorporates these ideas (in a more complex and refined fashion) when proposing the Information Load Hypothesis [ILH]. The ILH broadly establishes that increased processing cost has to be justified by additional discourse functions. For the ILH, the cost is a product of the conceptual representation (based on the “semantic” distance between the anaphor and the representation of the antecedent) and of the anaphoric form. The discourse function is related to both identifying the referent and adding new information. The mapping process is given by a direct relation between cost and discourse function: the less specific the representation of the anaphor is with respect to the representation of the antecedent, the less costly the anaphor is to process. The acceptability of a given anaphor in a certain context is not a consequence of its formal class but rather of the relation between its cost and its discourse function. This approach makes perfect sense with respect to our experimental and corpus data. What seems to govern the distribution and interpretation of null and overt subjects is a “local specialization”: speakers try to optimize contrast taking into account not only *in presentia* elements but also the paradigmatic possible structures that are available in the language yet not used in a given context (following Baumann *et al.*, 2014; de la Fuente *et al.*, 2016; among others). Given the higher cost of the relation between the easily accessible antecedents and the overt subjects in Experiment 2, null subjects are preferred. In Experiment 1, the competition between the antecedents justifies the use of an overt subject, a costly solution, to avoid the ambiguity in the context. Recently, Almor *et al.* (2017) also show the effect of the ILH in the use of third-person anaphora in BP. Future research will be necessary to investigate up to which point the level of Complexity in the use of first-person singular overt and null subjects associated to explicitly marked verb forms is a function of the reversed mapping with the cost of identifying the referent.

REFERENCES

ALMOR, A. *Noun-phrase anaphora and focus: the informational Load Hypothesis*. 1996. Thesis (PhD in Cognitive Science) – Brown University, Los Angeles, CA, USA, 1996.

ALMOR, A. Noun-phrase anaphora and focus: the informational Load Hypothesis. *Psychological Review*, v. 106, n. 4, p. 748–765. 1999.

_____. Constraints and mechanisms in theories of anaphor processing. In: CROCKER, M. W.; PICKERING, M.; CLIFTON, C. (ed.). *Architectures and mechanisms for language processing*. Cambridge University Press, 2000. p. 341–354.

ALMOR, A.; DE CARVALHO MAIA, J.; CUNHA LIMA, M.; VERNICE, M.; GELORMINI-LEZAMA, C. Language processing, acceptability, and statistical distribution: a study of null and overt subjects in Brazilian Portuguese. *Journal of Memory and Language*, v. 92, 98–113, 2017.

ARIEL, M. *Accessing noun-phrase antecedents*. Routledge, London, UK, 1990.

_____. Interpreting anaphoric expressions: a cognitive versus a pragmatic approach. *Journal of Linguistics*, v. 30, p. 3–42, 1994.

_____. Accessibility theory: an overview, In: SANDERS, T.; SCHLIPEROORD, J.; SPOOREN, W. (ed.) *Text representation*, Springer Verlag, 2001. p. 29–87.

ARNOLD, J. E. *Reference form and discourse patterns*. 1998. Thesis (PhD in Linguistics) – Stanford University, Stanford, CA, USA, 1998.

BAAAYEN, R. H., DAVIDSON, D. J.; BATES, D. M. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, p. 390–412, 2008.

BARBOSA, M. d. P. P. *Partial pro drop as Null NP Anaphora*. Ms., University of Lisbon, 2011.

BARBOSA, M. D. P. P.; DUARTE, M. E. L.; KATO, M. A. null subjects in European and Brazilian Portuguese, *Journal of Portuguese Linguistics*, v. 4, n. 2, p. 11–52, 2005.

BARR, D. J.; LEVY, R.; SCHEEPERS, C.; TILY, H. J. Random effects structure for confirmatory hypothesis testing: keep it maximal, *Journal of Memory and Language*. v. 68, n. 3, p. 255–278, 2013.

BATES, D.; MAECHLER, M. *Lme4: linear mixed-effects models using S4 classes*, R package version 0.999375-27, 2009.

BATES, D.; MAECHLER, M.; BOLKER, B.; WALKER, S. Fitting linear mixed-effects models using lme4, *Journal of Statistical Software*, v. 67, n. 1, 1–48, 2015.

BAUMANN, P.; KONIECZNY, L.; HEMFORTH, B. Conversational implicatures in anaphora resolution: alternative constructions and referring expressions. In: HEMFORTH, B.; MERTINS, B.; FABRICIUS-HANSEN, C. (ed.). *Psycholinguistic Approaches to Meaning and Understanding Across Languages*, Springer Verlag, 2014. p. 197-212.

BIBERAUER, T., HOLMBERG, A., ROBERTS, I.; SHEEHAN, M. *Parametric variation: null subjects in minimalist theory*. Cambridge University Press, Cambridge, MA, USA, 2010.

BICK, E. *The parsing system “Palavras”: Automatic grammatical analysis of Portuguese in a constraint grammar framework*, University of Oklahoma Press, Aarhus, Denmark, 2000.

CARMINATI, M. N. *The processing of Italian subject pronouns*. 2002. Thesis (PhD in Linguistics) – University of Massachusetts, Cambridge, MA, USA, 2002.

CHOMSKY, N. *Lectures in government and binding*. Foris, Dordrecht: The Netherlands, 1981.

DE LA FUENTE, I. *Putting pronoun resolution in context: the role of syntax, semantics, and pragmatics in pronoun interpretation*. 2015. Thesis (PhD in Theoretical, Computational and Descriptive Linguistics) – University Sorbonne Paris City, Paris, France, 2015.

DE LA FUENTE, I., HEMFORTH, B., COLONNA, S.; SCHIMKE, S. The role of syntax, semantics, and pragmatics in pronoun resolution: a cross-linguistic overview, *In: HOLLER, A.; SUCKOW, K. (ed.). Empirical perspectives on anaphora resolution: information structural evidence in the race for salience*. Mouton de Gruyter, 2016. p. 11-32.

DRUMMOND, A. *Ibex farm*. Online server: <http://spellout.net/ibexfarm>. 2014.

DUARTE, M. E. L. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. *In: ROBERTS, I.; KATO, M. A. (ed.). Português brasileiro: uma viagem diacrônica*. Homenagem a Fernando Tarallo. Editora da Unicamp, Campinas, Brazil, 1993. p. 107–129.

DUARTE, M. E. L. *A perda do princípio “evite pronome” no português brasileiro*. 1995. Tese (Doutorado) – Universidade Estadual de Campinas, Campinas, SP, Brazil, 1995.

DUARTE, M. E. L. The loss of the ‘avoid pronoun’ principle in Brazilian Portuguese. *In: KATO, M. A.; NEGRÃO, E. V. (ed.). Frankfurt: Vervuert-Iberoamericana, 2000. p. 17–36.*

DUARTE, M. E. L. Avanço no estudo da mudança sintática associando a teoria da variação e mudança e a teoria de Princípios e Parâmetros. *Cadernos de Estudos Lingüísticos*, v. 57, n. 1, p. 85-111, 2015.

FERNANDES, E. G.; LUEGI, P.; CORREA SOARES, E.; DE LA FUENTE, I.; HEMFORTH, B. Adaptation in pronoun resolution: evidence from Brazilian and European Portuguese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 44, n. 12, p. 1986-2008, 2018.

FERNANDEZ-SORIANO, O. Strong pronouns in null subject languages and the Avoid Pronoun Principle. *MIT Working Papers in Linguistics*, v. 11, p. 228–240, 1989.

FERREIRA, M. B. Hyperraising and Null Subjects in Brazilian Portuguese. *MIT Working Papers in Linguistics: Collected Papers on Romance Syntax*, v. 47, p. 57–85, 2004.

FIGUEIREDO-SILVA, M. C. Main and embedded null subjects in Brazilian Portuguese. *In: KATO, M. A.; NEGRÃO, E. V. (ed.).*

Frankfurt: Vervuert-Iberoamericana, 2000. p. 17–36.

GILLIGAN, G. M. *A cross-linguistic approach to the pro-drop parameter*. 1987. PhD (PhD thesis) – University of Southern California, San Diego, CA, USA, 1987.

GIVÓN, T. Topic, pronoun and grammatical agreement. In: LI, C. N. (ed.). *Subject and Topic*, v. 6, Academic Press, New York, USA, 1976. p. 151-188.

GIVÓN, T. *Topic continuity in discourse - a quantitative cross language study*. John Benjamins, Amsterdam, The Netherlands, Philadelphia, USA, 1983.

GRIES, S. T. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, v. 10, n. 1, 2015. p. 95-125.

GUNDEL, J.; HEDBERG, N.; ZACHARSKI, R. Cognitive status and the form of referring expressions in discourse. *Language and Cognitive Processes*, v. 69, p. 274–307, 1993.

HAWKINS, J. A. *Efficiency and complexity in grammars*. Oxford University Press, Oxford, UK, 2004.

HAWKINS, J. A. *Cross-linguistic variation and efficiency*. Oxford University Press, Oxford, UK, 2014.

HOLMBERG, A., NAYUDU, A.; SHEEHAN, M. Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish, and Marathi. *Studia Linguistica*, p. 59–97, 2009.

JAEGGLI, O.; SAFIR, K. *The null subject parameter*. Kluwer, Dordrecht, The Netherlands, 1989.

KATO, M. A. Strong and weak pronominals in the null subject Parameter. *Probus*, v. 11, n. 1, p. 1–37, 1999.

KATO, M. A.; NEGRÃO, E. V. *Brazilian Portuguese and the null subject parameter*. Vervuert-Iberoamericana, Frankfurt, Germany, 2000.

LEE, J. W.; JONES, P. S.; MINEYAMA, Y.; ZHANG, X. E. Cultural differences in responses to a likert scales. *Research in Nursing & Health*, p. 295-306, 2002.

MODESTO, M. Null Subjects without “rich” agreement. In: M. A. KATO; E. V. NEGRÃO. (ed.). Frankfurt: Vervuert-Iberoamericana. 2000. p. 147–174.

MODESTO, M. Null Subjects in Brazilian Portuguese and Finnish: They are not derived by movement. In: DAVIES, W. D.; DUBINSKY, S. (ed.). *New horizons in the analysis of control and raising, studies in natural language and linguistic theory*. v. 71, Springer, Amsterdam, The Netherlands, 2008a. p. 231–248.

_____. Topic prominence and null subjects. In: BIBERAUER, T. (ed.). *The limits of syntactic variation, linguistics today*. Vol. 132, Amsterdam: John Benjamins, 2008b. p. 375–409.

NEGRÃO, E. V. *A distribuição e a interpretação de pronomes na fala de crianças da escola pública*. Ms., University of São Paulo, São Paulo, 1990.

NEGRÃO, E. V.; MÜLLER, A. L. D. P. As mudanças no sistema pronominal do português do Brasil: substituição ou especialização de formas?, *DELTA*, v. 12, p. 125–152, 1996.

NEWMAYER, F. J. Against a parameter-setting approach to language variation. *Linguistic Variation Yearbook*, v. 4, p. 181–234, 2004.

_____. *Possible and probable languages: a generative perspective on linguistic typology*. Oxford University Press, Oxford, UK, 2005.

_____. *A rejoinder to 'On the role of parameters in Universal Grammar: a reply to Newmeyer's by Ian Roberts and Anders Holmberg'*. Ms., revised version, February, 2006.

NUNES, J. Subespecificação de traços-f e hiperalçamento em português brasileiro. In: FIGUEIREDO, C.; ARAÚJO, E. (org.). *Diálogos com Ribeiro: Sobre gramática e história da língua portuguesa*. Salvador: Edufba, 2015. p. 121-148.

PERLMUTTER, D. *Deep and surface structure constraints in syntax*. Holt, Reinhart and Winston, New York, NJ, 1971.

RITTER, E. On the syntactic category of pronouns and agreement. *Natural Language and Linguistic Theory*, v. 13, n. 3, p. 405–443. 1995.

RIZZI, L. *Wh-movement, negation and the pro-drop parameter*. Ms., Scuola Normale Superiore, Pisa, Italy, 1980.

_____. *Issues in Italian syntax*. Foris, Dordrecht: The Netherlands, 1982.

_____. Null objects in Italian and the theory of pro. *Linguistic Inquiry*, v. 17, p. 501–557, 1986.

_____. *On the grammatical basis of language development: a case study*. Ms., Siena University, Siena, Italy, 2002.

ROBERTS, I. Taraldsen's Generalization and language change: two ways to lose null subjects. In: SVENONIUS, P. (ed.). *Functional structure from top to toe: the cartography of syntactic structures*. v. 9, Oxford University Press, Oxford: UK, 2014. p. 115-148.

ROBERTS, I. *The null subject parameter in the 21st century*, Ms. Downing College, University of Cambridge, Cambridge, UK, 2016.

ROBERTS, I.; HOLMBERG, A. Introduction. In: BIBERAUER, T.; HOLMBERG, A.; ROBERTS, I.; SHEEHAN, M. (ed.). Amsterdam: John Benjamins, 2010. p. 1–56.

ROHRBACHER, B. *The Germanic VO Languages and the Full Paradigm: A Theory of V to I Raising*. PhD thesis, University of Massachusetts, Cambridge, MA, USA, 1994.

SOARES, E. C. *Anaphors in discourse: null subjects in Brazilian Portuguese*. 2017. Thesis (PhD in Theoretical, Computational and Descriptive Linguistics)– University Sorbonne Paris City, Paris, France, 2017.

SHLONSKY, U. Hebrew as a partial null-subject language. *Studia Linguistica*, v. 63, n. 1, p. 133–157, 2009.

SIMONENKO, A.; CRABBÉ, B.; PRÉVOST, S. Taraldsen's generalization in diachrony: evidence from a diachronic corpus. In: KAPLAN, A. (ed.). *Proceedings of the 34th West Coast Conference on Formal Linguistics*, CSLI Publications, Somerville, MA, USA, 2017. p. 462–469.

TARALDSEN, K. T. *On the NIC, vacuous application and the that-trace filter*. Indiana University Linguistics Club, 1980.

TARALLO, F. *Relativization strategies in brazilian portuguese*. 1983. Thesis (PhD in Linguistics), University of Pennsylvania, 1983.

Acknowledgements

Research reported in this paper was mostly carried out while the first author was a PhD student at the University of Paris VII Denis Diderot supervised by the other two authors and Sergio Menuzzi. Over this period, this study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 (trial number: 0628/14-0). Some parts of the text first appeared in Soares (2017); the conclusions and crucial evidence (Experiment 2) presented in this paper had not been published before. The final version of the paper was written while the first author was holding a post-doc position in the Federal University of Santa Catarina in the National Program of Post-Doctorate (PNPD) of the CAPES (trial number: 88882.316438/2019-01). We are thankful to the labs where this study was developed (the CLILLAC-ARP and the LLF at University of Paris VII Denis Diderot), to the LabLing at UFSC, and to their members for their support. We are also thankful to the researchers who made comments on preliminary versions of this paper: Jeffrey Runner, Scott Schwenter, Sergio Menuzzi and Anne Abeillé, the audience in the National Meeting of the Working Group of Grammatical Theory (GT-TG – Anpoll) and two anonymous reviewers for FL. Finally, we are thankful to the participants who answered the online questionnaires.



Received in October 28, 2018. Approved in February 25, 2019.