

RESEARCH PAPER

# Subject Pronoun Expression in Santomean Portuguese

Marie-Eve Bouchard

New York University, US  
mebouchard@nyu.edu

---

Studies on Subject Pronoun Expression (SPE) in the Portuguese-speaking world have shown a distinction between European Portuguese, which is a Null Subject Language (NSL) with high rates of null subjects, and Brazilian Portuguese, which is controversially treated as a partial-NSL and exhibits a considerably lower rate of null subjects. No specific studies have been conducted on the matter on Santomean Portuguese, but we know that both null and overt subject personal pronouns exist in this variety of Portuguese. The objective of this paper is to investigate variation in SPE in Santomean Portuguese, and to situate this variety of Portuguese in comparison with other varieties. Results of the variationist analyses show that Santomean Portuguese patterns more like European Portuguese in its high rate of use of null subject. Interestingly, and contrary to previous studies, Santomeans with a higher level of education disfavor the use of null subject, which I relate to a sensitivity to grammatical ideology and the favoring of the overt subject in more formal situations. Most of the results regarding the linguistic predictors, which are stronger than the social predictors, relate Santomean Portuguese to other varieties of Portuguese, and to Spanish.

---

**Keywords:** Subject Pronoun Expression; Santomean Portuguese; variationist sociolinguistics; language variation; null subject languages

---

## 1. Introduction

Languages that do not require the presence of an overt subject personal pronoun (henceforth SPP) are called Null Subject Languages (henceforth NSL), or pro-drop languages, and the ones that ordinarily require the presence of an overt SPP are called non-Null Subject Languages (henceforth non-NSL), or non-pro-drop languages. Italian, Spanish, and European Portuguese are NSL, while French, English, and German are non-NSL. Following are examples of null and overt SPP, where we see that SPP are obligatory in English (1), and optional in Spanish (2):

- (1)
  - a. you bought a computer
  - b. \*bought a computer
  
- (2)
  - a. *tu compraste un ordenador*
  - b. *compraste um ordenador*

Although there is no agreement on this classification, some languages are also considered to be partial-NSL, such as Brazilian Portuguese, Finnish, and Marathi. Those languages allow null subjects under more restricted conditions than full-fledged NSL (Holmberg, Nayudu & Sheehan 2009).

Variation in Subject Pronoun Expression (henceforth SPE) is of interest to sociolinguists because the speaker has the option of expressing the SPP or omitting it. How does a speaker make a choice between those two options? The main objective of most sociolinguistic research on SPE has been to ascertain the linguistic, stylistic, and social factors that determine, or at least that influence, the expression or omission of the SPP. All sociolinguistic research has found correlations between those factors and the SPE. Even so, this syntactic variable remains highly debated among scholars who work on the topic.

Linguists have investigated SPE in Brazilian and European Portuguese (cf. Barbosa 2000, 2009; Barbosa, Duarte & Kato 2001, 2005; Duarte 1993), but no studies have been done on SPE in the variety of Portuguese spoken in São Tomé and Príncipe. All we know is that both null and overt SPP are present in Santomean Portuguese, as shown in (3):

- (3) STP: os Angolares, **eles** não falam peixe, [Ø] falam *kikiê*  
 ENG: the Angolar they NEG say.3PL fish, speak.3PL *kikiê*  
 ‘the Angolares, they don’t say fish, they say *kikiê*’  
 - Suéli, 32 years old

My objective is to investigate the social and linguistic factors that condition SPE in Santomean Portuguese, and to compare the results to previous research on SPE in Brazilian and European Portuguese.

São Tomé and Príncipe was a Portuguese colony from the late fifteenth century to 1975. From the sixteenth century to the beginning of the twentieth century, Forro, Angolar, and Lung’ie (three native creoles) were the most widely spoken languages on the islands (Hagemeyer, in press). However, the massive arrivals of contract laborers starting at the end of the nineteenth century, and the use of Portuguese as a lingua franca completely changed the sociolinguistic setting. As a consequence, a process of linguistic shift (from creoles to Portuguese) started to take place. This shift was intensified from the 1960s, with the rise of the nationalist movement, the choice of Portuguese as a unifying language for Santomeans of different ethnic background, the independence of the country (in 1975), and the generalized access to education (Bouchard 2017; Seibert 2006). Since then, children have been growing up with the local variety of Portuguese as their first (and often only) language (Bouchard 2017). Today, 98.4% of the population speak Portuguese (as a first or second language) (INE 2012).

This paper’s first section provides a background on SPE in Portuguese varieties. The second section details the methodology for analyzing SPE and the social and linguistic variables included in the quantitative analyses. The third section offers an overview of the distribution of null and overt subject pronouns in Santomean Portuguese, and the fourth section presents and analyzes the results. Finally, the last section is a wrap-up of the most important findings.

## 2. Subject Pronoun Expression in Portuguese varieties

Variable SPE constitutes a morphosyntactic feature that Portuguese inherited from Latin. However, not all Latin’s descendant languages developed into NSL: European Portuguese, Spanish, and Italian are still consistent NSL, Brazilian Portuguese is a partial-NSL, or a NSL with a lower rate of null subjects, depending on one’s position, and Modern French and Haitian Creole are non-NSL (Orozco 2015a). This section will review some of the literature on SPE in varieties of Portuguese, focusing on European and Brazilian Portuguese, as more studies exist on those varieties, in order to present how speakers of Portuguese use SPP.

The nature of null subjects in Brazilian Portuguese has been investigated by Negrão (1997), Modesto (2000a, 2000b, 2009), Rodrigues (2002, 2004), and Sheehan (2006), and in European Portuguese by Duarte (1995) and Barbosa (1995, 2000, 2009). Studies comparing SPE in European and Brazilian Portuguese are also numerous (cf. Barbosa, Duarte & Kato 2001; Magalhães & Santos 2006). As written above, European Portuguese is a full-fledged NSL, while Brazilian Portuguese is undergoing language change in the direction of becoming a non-NSL (Martínez-Sanz 2011); it is considered a partial-NSL (cf. Holmberg, Nayadu & Sheenan 2009), or a semi pro-drop language (cf. Erker & Guy 2012). In one view, this partiality implies that “two different NSL grammars are available in the mental grammars of Brazilian speakers: on the one hand, a NSL grammar that allows for the licensing of null subjects, and a non-NSL grammar, on the other, responsible for widespread overt subjects” (Martínez-Sanz 2011: 154–155). The Brazilian SPE system could be viewed as a language that allows null subjects in certain restricted environments, but that lacks unrestricted referential null subjects (Barbosa 2013; Martínez-Sanz 2011).

Lobo (2016) discusses variation in SPE comparing, among other languages, European and Brazilian Portuguese. According to her, one of the distinctions between these two varieties of Portuguese is the interpretation of the (null or overt) subject in reference to its antecedent. In European Portuguese, a null SPP usually refers to the subject of the main sentence, while an overt SPP usually refers to the complement. That means that in (4), it is João who won the race, and in (5), it is Pedro.

- (4) EP: o João disse ao Pedro que  $\emptyset$  tinha ganho a corrida  
 ENG: the John told.3SG to.the Peter that had.3SG won the race  
 ‘John<sub>i</sub> told Peter<sub>j</sub> that he<sub>i</sub> won the race’  
 Lobo (2016: 564)

- (5) EP: o João disse ao Pedro que **ele** tinha ganho a corrida  
 ENG: the John told.3SG to.the Peter that he had.3SG won the race  
 ‘John<sub>i</sub> told Peter<sub>j</sub> that he<sub>j</sub> won the race’  
 Lobo (2016: 564)

In Brazilian Portuguese (a partial-NSL), the overt SPP is not necessarily interpreted the same way as in European Portuguese, as the overt SPP in (6) may refer to João, or to Pedro.

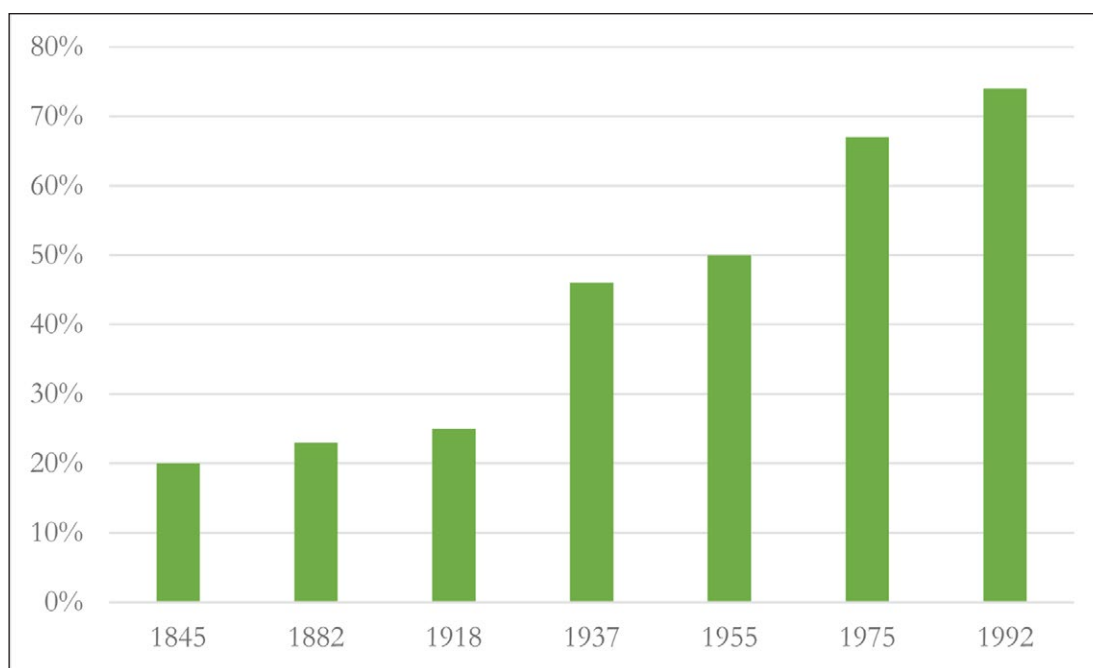
- (6) BP: o João disse ao Pedro que **ele** tinha ganho a corrida  
 ENG: the John told.3SG to.the Peter that he had.3SG won the race  
 ‘John<sub>i</sub> told Peter<sub>j</sub> that he<sub>i/j</sub> won the race’

That being said, the pioneer work on the changing nature of SPE in Brazilian Portuguese is Duarte’s (1995) dissertation. In her study, she demonstrates how Brazilian Portuguese differs from European Portuguese and other NSL regarding SPE, and how it is changing toward a more frequent use of overt subjects. **Figure 1** illustrates this change.

The author relates this change in SPE to the reduction of the Brazilian inflectional paradigm. This reduction probably started with the loss of the second person singular *tu* ‘you’ and its replacement by *você* (which takes third person agreement) (Duarte 1993; Galves 1990). From the inflectional paradigm that originally showed six distinctive forms, Brazilian Portuguese now has three distinctive verb endings, as a result of the loss of *tu*

‘you’ and the replacement of first person plural *nós* ‘we’ by the expression *a gente* ‘we’, which also takes third person singular verbal agreement (e.g. *nós comemos* → *a gente come* ‘we eat’).<sup>1</sup> To illustrate this change in the Brazilian inflectional paradigm, see the difference in **Table 1** between the European and Brazilian systems with the verb *falar* ‘to speak/to talk’.

According to Duarte (1995), as a consequence of this changing paradigm, the Avoid Pronoun Principle<sup>2</sup> that usually leads to the null representation of the subject is lost, and null subject becomes a less frequently used option. In her work, she showed that in the 1992 stage, 71% of SPP were phonologically realized, and 29% were not. However, this is a functional explanation of the change in Brazilian Portuguese. An alternative theory is that reduced verbal inflection and higher rates of SPE are both consequences of slavery in Brazil, and massive L2 acquisition of some perhaps creolized but certainly non-standard version of Portuguese by the Africans brought to Brazil (Guy 1981). A semi-acquired L2 version of Portuguese (as well as a creole) would probably lack verbal inflection and require overt SPP. In this view, contemporary Popular Brazilian Portuguese is a partially decreolized descendant of that earlier L2 version of the language (cf. Guy 1981; Lucchesi et al. 2009). However, it is somewhat reductionist to only include African descendants in this theory, as other linguistic and cultural groups (including Amerindians and European emigrants) also acquired Portuguese as a L2 in Brazil. Also, under a decreolization hypothesis, one might expect the language to go from non-NSL to NSL, i.e. the opposite of what is shown in **Figure 1**.



**Figure 1:** Overt pronominal subjects in BP through seven periods.<sup>3</sup> (Barbosa, Duarte & Kato 2005: 6; adapted from Duarte 1993).

<sup>1</sup> The third person plural SPP *nós* ‘we’ has not totally disappeared in Brazil, and some regions use it more than others, but Duarte (1993) argues that *nós* is mainly used in writing, and in spoken language by an older generation.

<sup>2</sup> This is a principle from Chomsky (1981) that states “Avoid Pronoun”, and imposes a choice of a null subject over an overt subject when possible. It is a subcategory of another conversational principle that states “Don’t say more than is required”.

<sup>3</sup> The corpora analyzed between 1845 and 1992 come from popular plays.

Note that other studies on the syntax of subject licensing in Brazilian Portuguese agree with Duarte (1995) regarding the semantic and syntactic distinction of null subjects in this language (Barbosa 2009, 2013; Ferreira 2000; Holmberg 2005; Kato 1999; Rodrigues 2002, 2004; Sheehan 2006), which set it apart from European Portuguese and other NSL.

There are not as many quantitative studies on SPE in European Portuguese as there are in Brazilian Portuguese.<sup>5</sup> Much of the information we have about SPE in European Portuguese come from comparative studies between this language and Brazilian Portuguese.

Duarte (2000) has compared the distribution of overt and null subjects in those varieties of Portuguese. In **Table 2** (adapted from Barbosa, Duarte & Kato 2005: 23; and Duarte 2000: 25) are the results of her study, which was based on spoken corpora.

This table shows that European Portuguese favors null subjects and that Brazilian Portuguese favors overt subjects. Those numbers vary depending on the person; in **Figure 2**, we see that in Brazilian Portuguese overt subjects occur with the greatest frequency with second person while in European Portuguese they do so with first person.

All the studies mentioned above have demonstrated the difference between European and Brazilian Portuguese regarding SPE.

Very little work has been done on African varieties of Portuguese regarding SPE. Oliveira and Ferreira dos Santos (2007) and Teixeira (2013) investigated SPE in Angolan Portuguese. First, Oliveira and Ferreira dos Santos (2007) examined the pronominal system of Angolan Portuguese and noted how it is becoming more like Brazilian Portuguese in relation to the use of *você* ‘you’ and *a gente* ‘we’. However, as is the case in European Portuguese, Angolan speakers of Portuguese also use *tu* ‘you’; there is therefore variation between the two second person singular forms. **Table 3** (adapted from Santos 2006, in Oliveira & Ferreira dos Santos 2007: 97) shows their results regarding SPE.

We note in **Table 3** that 1) like European and Brazilian Portuguese, both overt and null subjects are possible, and 2) that Angolan Portuguese seems to pattern more like European

**Table 1:** Spoken Brazilian and European inflectional paradigms.

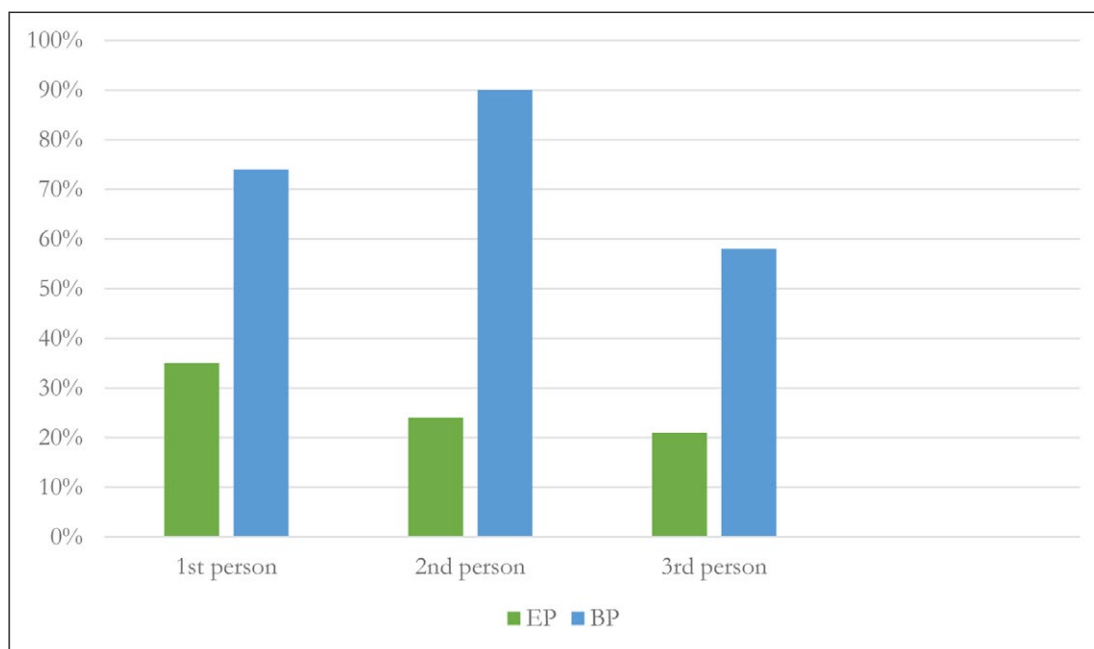
PERSON & NUMBER	BRAZILIAN PORTUGUESE	EUROPEAN PORTUGUESE
1sg	(eu) falo	(eu) falo
2sg <sup>4</sup>	—	(tu) falas
3sg	(você, ele/a, a gente) fala	(você, ele/a) fala
1pl	—	(nós) falamos
2pl	—	(vós) falais
3pl	(vocês, eles/as) falam	(vocês, eles/as) falam

**Table 2:** Percentages of null and overt subjects in EP and BP.

Variety	Null subjects	Overt subjects
EP	73.3%	26.7%
BP	26.0%	74.0%

<sup>4</sup> Making generalizations about Brazilian Portuguese is challenging, as many different dialects of the variety exist. Regarding the second person singular for example, note that in some regions such as Rio Grande do Sul and Pará, the pronoun *tu* ‘you’ is used, but it usually agrees with the third person singular verb ending (e.g. *tu fala* instead of *tu falas* ‘you talk’).

<sup>5</sup> One may refer to Ambulate (2008), and Costa, Lobo and Silva (2009), for work on acquisition of SPE, and to Barbosa (2000, 2010, 2013), Lobo (1995), and Raposo (1986) for work on the syntax of SPE.



**Figure 2:** Overt subjects in spoken EP and BP. (Barbosa, Duarte & Kato 2005: 22; adapted from Duarte 2000: 25).

**Table 3:** Frequency number of SPP in Angolan Portuguese (369 sentences).<sup>6</sup>

SPP	OVERT		NULL	
	#	%	#	%
<i>Eu</i>	87/285	30.5	198/285	69.5
<i>Tu</i>	1/18	5.5	17/18	94.5
<i>Você</i>	1/141	0.7	140/141	99.3
<i>Ele/Ela</i>	10/21	47.6	11/21	52.4
<i>Nós</i>	38/97	39.2	59/97	60.8
<i>A gente</i>	1/1	100.0	--	0.0
<i>Vocês</i>	7/20	35.0	13/20	65.0
<i>Eles/Elas</i>	2/8	25.0	6/8	75.0
<b>TOTAL:</b>	147/591	24.9	444/591	75.1

Portuguese and shows a higher number of null subjects (75.1%) than overt subjects (24.9%). Oliveira and Ferreira dos Santos highlight that the numbers of overt pronouns in the first persons are higher than in the other persons: 30.5% for first person singular (*eu*), and 39.2% for first person plural (*nós*) (with only one token for the other form of first person plural, *a gente*). These numbers show similarity with European Portuguese, where overt SPP is used in 35% of the total occurrences in the first person singular, for example (Figure 2). But Angolan Portuguese behaves differently from Brazilian Portuguese, with a very low number of overt SPP in the second person singular (5.5% for *tu*, and 0.7% for *você*). Note that Oliveira and Ferreira dos Santos do not discuss the results for the third person singular, although the use of overt SPP is high (47.6%).

<sup>6</sup> Note that Oliveira and Ferreira dos Santos (2007) took their data from Santos (2006), who reorganized data from Chavagne's (2005) dissertation. Chavagne (2005) wrote an exhaustive description of Angolan Portuguese.



Teixeira's (2013) variationist study of Angolan Portuguese gives different results; according to his data, 65% of SPP are overt, and 35% are null. Those results are very similar to the ones found by Duarte (1995) for Brazilian Portuguese. However, not enough information about the methodology is given in Oliveira and Ferreira dos Santos (2007) and Teixeira (2013) to analyze and understand the difference between their results.

To my knowledge, Dias (2009) is the only paper on SPE in Mozambican Portuguese. However, her study differs greatly from the other mentioned above as she used a written corpus. Dias investigated SPE in the written Portuguese of forty-five 5th grade students in a suburban region of Maputo. They are all bilingual, speakers of Changana and Portuguese, and most of them learned Portuguese as a L2 at school. Her results show that 52.5% are null SPP and 47.5% are overt. Interestingly, Dias noted that the first person singular mainly occurs with overt SPP, while the first person plural only occurs with null SPP. She also writes that null SPP use correlates with more verbal agreement.

In **Table 4** are results that compare numbers for the four varieties of Portuguese discussed in this section. However, remember that the first three varieties are in their spoken form, and the last one, in written form. These numbers come from different studies and are not balanced regarding sociolinguistic variables. More comparable studies on the topic are necessary.

To my knowledge, there are no studies on SPE in the variety of Portuguese spoken in São Tomé and Príncipe. However, it is possible to see in the literature on Santomean Portuguese that SPPs can be expressed (7) or not (8):

- (7) Gonçalves (2010: 130)  
 STP: **nós** fazemos um pouco de tudo sem aprofundarmos bem no assunto  
 Eng: we do.3PL a little of all without deepening.3PL well in.the topic  
 'we do a little bit of everything without deepening too much in any topic'
- (8) Gonçalves (2010: 50)  
 STP: queria matricular no instituto  
 Eng: wanted.1SG to.register in.the institute  
 'I wanted to register for the institute'

It is also possible to find sentences that contain both overt and null SPP, as in example (9):

- (9) Gonçalves (2010: 130)  
 STP: depois cheguei (a) um momento que **eu** vi que era vazio  
 Eng: after arrived.1SG (to) one moment that I saw.1SG that was.3SG empty  
 'after, I arrived at some point and saw that it was empty'

**Table 4:** SPE in European, Angolan, Brazilian, and Mozambican Portuguese.

Variety	Null subjects	Overt subjects
EP	73.3%	26.7%
AP <sup>7</sup>	75.1%	24.9%
BP	26.0%	74.0%
MP	52.5%	47.5%

<sup>7</sup> Oliveira & Ferreira dos Santos (2007). Note that Teixeira (2013) had slightly different results, with 65% of overt SPPs.

Based on this information, I now turn to the methodology of the variationist analysis I conducted to investigate SPE in Santomean Portuguese.

### 3. Methodology for coding

The fieldwork for my data was mainly conducted in the city of São Tomé, the capital of São Tomé and Príncipe, and its surroundings, between June 2015 and March 2017. This study is based on roughly 46 hours of tape-recorded individual interviews from 48 adults and eight teenagers. These interviews were carried out employing techniques from both sociolinguistic interviews (Becker 2013; Labov 1984; Tagliamonte 2006) and ethnographic interviews (Spradley 1979). The participants included in this study are Santomeans born and raised on São Tomé Island who live in the capital or its surroundings. Many of the participants are monolingual Portuguese speakers, or have some knowledge of Forro, and a few (usually older) participants are bilingual native speakers of Forro and Portuguese. The interviews were transcribed, and 100 tokens per participant were coded for analysis. Decisions regarding inclusion and exclusion of tokens are greatly influenced by the coding manual of Otheguy and Zentella (2012). The tokens included in the dataset are the ones where the null and overt subject alternation is possible. The initial dataset comprised 5,600 tokens (100 tokens per speaker), and was reduced to 4,512 tokens once the full noun phrase subjects were excluded. The envelope of variation is based around the Principle of Accountability (Labov 1972), i.e. all clauses where the variant is possible are analyzed to compare the number of tokens of null subjects with those of expressed subjects.

#### 3.1. Dependent linguistic variable

TYPE OF PRONOUN EXPRESSION. The dependent variable is how speakers express a SSP, whether it is with a null subject (e.g. *falas* ‘you speak’) or an overt pronominal subject (e.g. *tu falas* ‘you speak’). This is summarized in **Table 5**.

#### 3.2. Independent social variables

SPEAKER. I have chosen 56 speakers from the capital of São Tomé and its surroundings. Region is a social variable that has been widely discussed in language variation and dialect studies (cf. Chambers & Trudgill 1998). The place where people grew up and spent most of their time is traditionally an important criterion when studying variation; speakers from different places speak differently. To be rigorous about this, all the participants of this study come from the same area.

GENDER. I selected an equal number of men and women in order to study the variable gender. Sociolinguistic studies have shown that linguistic variation often correlates with gender (or sex) of speakers (Cheshire 2004; Trudgill 2000). Following Eckert (1990), I choose the word *gender* to refer to the social and cultural elaboration of sex difference, as *sex* has become more politicized and problematized in the past few decades (Cheshire 2004). Gender separation is manifested in a number of domains of social life in São Tomé, including the division of labor regarding housekeeping, parenting, tasks and jobs, among other things.

AGE. The speakers selected can be divided into five age categories: 12–18, 20–29, 30–39, 40–49, and 50 and more. Note that São Tomé and Príncipe has a youthful age

**Table 5:** Dependent linguistic variable.

Variable	Variants
subject personal pronoun expression	null subject overt subject



structure, with 60% of the population under the age of 25, and 6.5% over 55 (CIA World Factbook 2017). This explains why the age categories are low, and why I do not have more categories for older people. The correlation of linguistic variables with age is important when studying language change; in this case, investigating age will allow us to investigate change in progress (Labov 1963, 1966), applying the apparent-time construct (Bailey 2004; Bailey et al. 1991).

**LEVEL OF EDUCATION.** Level of education is a good indicator of socioeconomic status in São Tomé, as in many other countries. Many sociolinguistic studies provide evidence that different social groups within a community differ in their usage of linguistic features (cf. Labov 1966; Trudgill 1974). For this study, level of education is divided into primary school (grade 1 to 6), high school (grade 7 to 12), and university (including bachelor, master and doctorate). All participants attended school, and some of them were still in school at the time of the interviews. The school grade that was attributed to them is the grade that was completed, or in progress in the case of those who were still in school.

**ETHNIC ORIGIN.** Labelling by ethnic origin is problematic (cf. Fought 2004), especially among a mixed-race and mixed-ethnic population such as São Tomé and Príncipe. I tried as much as possible to focus this research on Forros because, as for place of origin within the island, I am not sure if ethnic origin has an influence on language or not. By choosing mainly Forros to participate to this study, I wanted to avoid dealing with this problem. However, this did not work out as planned, as some of my “Forro” participants appeared to have one non-Forro parent (Angolar or Cape Verdean). Therefore, participants are divided into two (unbalanced) groups, depending on if they have two Forro parents, or one Forro parent (the other one being of other ethnic origin).<sup>8</sup>

**SPOKEN LANGUAGE(S).** The possible influence of creole on Portuguese is important to the present study. All the participants speak Portuguese, but knowledge of creole varies from one speaker to another. I divided speakers according to whether they were monolingual in Portuguese L1 (with no knowledge of creole), speakers of Portuguese L1 with “some” knowledge of creole, and bilingual.<sup>9</sup>

**Table 6** summarizes the independent social factors.<sup>10</sup>

### 3.3. Independent linguistic variables

**TYPE OF CLAUSE.** To see whether the type of clause has an impact on SPE, the clauses were divided into three groups: main clause, conjoined clause, and subordinate clause. **TYPE OF CLAUSE** appeared as a constraint that significantly conditions SPE in previous studies of Spanish (Morales 1997; Orozco 2015a; Otheguy & Zentella 2012; Otheguy, Zentella & Livert 2007).

**PRIMING EFFECTS.** Priming is a psycholinguistic process that consists of the repetition of an element or linguistic structure (Cameron & Flores-Ferrán 2004; Flores-Ferrán 2002; Travis 2005). Understanding the role of priming in speech can impact our understanding of SPE. This variable is divided into four groups: previous clause had a full NP subject, previous clause had an overt SPP, previous clause had a null SPP, and no priming. The “no priming” was usually used when the previous clause was not spoken by the interviewee,

<sup>8</sup> Forty-six participants have two Forro parents (representing 82% of the sample), and ten have only one Forro parent (representing 18% of the sample).

<sup>9</sup> The bilingual speakers with creole as L1 and Portuguese as L2, Portuguese as L1 and creole as L2, and the ones with both creole and Portuguese as L1 were put together in the same category, because they were not numerous.

<sup>10</sup> Three social categorizations were used to find participants: age, gender, and level of education. The adult sample (people over 20 years old in this case) are evenly distributed for gender, age and education level. What makes the entire sample unevenly distributed are the teenagers, as they are less numerous (eight), seven out of eight are in high school, and none of them had started at university.

or when there was a long pause or laughs. The hypothesis is variant continuity: use of one subject type favors the subsequent use of the same type (Orozco 2015a).

**MORPHOLOGICAL REGULARITY.** Verb forms were divided into regular and irregular verbs. The website *Conjuga-me* ([www.conjuga-me.net](http://www.conjuga-me.net)) was used to verify the morphological regularity of verbs. Verb forms were coded as irregular if there was a change in the root of the verb (e.g. *medir* ‘to measure’, *meço* ‘I measure’ and not \**medo*), and if there was a change in the regular ending of verb (e.g. *querer* ‘to want’, *ele quer* ‘he wants’ and not \**ele quere*). Each form was coded independently of the other forms of the same word; for instance, *leio* ‘I read’ was coded as irregular and *lemos* ‘we read’, as regular, even if they have the same root *ler* ‘to read’. Results for Spanish SPE in Erker and Guy (2012) show that irregular verbs are more often used with null subjects than regular verbs are, most notably among high-frequency verbs.

**SEMANTIC CONTENT.** Following Erker and Guy (2012), verb forms were divided into the following three semantic classes: mental activity (e.g. *saber* ‘to know’, *pensar* ‘to think’), stative (e.g. *ser/estar* ‘to be’, *ter* ‘to have’), and external activity (e.g. *correr* ‘to run’, *beber* ‘to drink’). Previous studies have shown that mental activity verbs have the lowest null subject rates and that external activity verbs have the highest (Erker & Guy 2012; Orozco 2015a).

**VERB FORM.** This contrasts complex verb forms (e.g. *tinha falado* ‘I had talked’, *vou dizer* ‘I’ll say’), and simple verbs (e.g. *falei* ‘I talked’, *digo* ‘I say’). Contrary to Otheguy and Zentella (2012), I did not consider *querer* ‘to want’ + infinitive to be a complex verb form. When a token is a complex verb, I look at the finite form of the verb to code it. For instance, for the token *vou passar* ‘I’ll pass’, *vou* is irregular, so the token is coded as irregular, even if *passar* is a regular verb.

**AMBIGUITY PARADIGM.** This refers to how clearly the verb form indicates what the subject is, based on its morphology. Some verb tenses provide a more obvious indication of the subject person/number (present, past tense, future, imperative) because all persons have a different ending. In others, this is less obvious (imperfect tense, conditional, subjunctive) because some persons have identical inflections. The idea behind this coding

**Table 6:** Independent social factors for coding SPE.

Factors	Levels	# of speakers
speaker	56 speakers	56
age	12–18	8
	20–29	12
	30–39	12
	40–49	12
	50 and more	12
gender	male	28
	female	28
level of education	primary school	17
	high school	23
	university	16
ethnic origin	two parents are Forros	46
	only one parent is Forro	10
spoken languages	monolingual (Portuguese L1)	12
	Portuguese L1, some creole L2	27
	bilingual (Portuguese and creole)	14

is that when the morphology of a verb makes it clear what person and number the subject is, an overt SPP may appear redundant.<sup>11</sup> Consequently, I expect that the verbs that have a “more obvious” morphology favor the use of null subject.

**PERSON AND NUMBER.** Verb forms were classified for one of the five person and number values: first singular, second singular, third singular, first plural, and third plural. Note that there were no tokens for second person plural, and the over SPP *vós* ‘you.2PL’ is an old form that is basically no longer used in spoken Portuguese. The tokens for PERSON AND NUMBER are unevenly distributed, with 1,595 first singular, 46 second singular, 2,037 third singular, 406 first plural, and 428 third plural. When there was absence of verbal agreement (e.g. *tu fala* ‘you speak.3SG’), it is still the verb form that was coded, and not the subject – therefore, *tu fala* is coded as third person singular. Based on Duarte (1993) who has illustrated the increased use of overt subjects over time (1845–1992) in Brazilian Portuguese, the difference between each person and number might not be as significant as the changes across time (**Figure 1**).

**ANIMACY.** Animacy is expressed based on how “alive” the referent of a noun is, whether it is animate or inanimate. Barbosa, Duarte and Kato (2005) have shown that null subjects in European Portuguese are strongly favored when the subject referent of a verb is inanimate.

**COREFERENTIALITY.** Coreferentiality is defined in terms of the relationship between the target verb (i.e. the one being examined) and the trigger verb (i.e. the finite verb preceding the target verb). It is also sometimes referred to as switch reference (e.g. Erker & Guy 2012). The present study made a distinction between coreference with subject of previous clause, coreference with indirect object complement (IO), coreference with direct object complement (DO), coreference with oblique object complement (OO), and switch reference. Switch reference has always been found to favor overt subjects.

**FREQUENCY.** To find the frequency of a word, I used the *Corpus do Português* (www.corpusdoportugues.org), and their corpus called *Web/Dialects*, which has 1 billion words (Davies & Ferreira 2017).<sup>12</sup> The different forms of a verb were considered all together; for instance, *como* ‘I eat’, *comeram* ‘they ate’, and *comemos* ‘we ate’ were all coded with the value 112 339 ( $\log(10) = 5$ ) under *COMER* ‘to eat’. The frequency of all tokens was then grouped into seven categories based on a logarithm (base 10). I expect high frequency tokens to behave differently than the low frequency ones (Erker & Guy 2012).

**Table 7** is a summary of these factors.

### 3.4. Considerations for coding

Coding for SPE was subject to the following considerations:

- 1) I only included finite verbs, and therefore did not include non-finite verbs (infinitive, participles, and gerunds). That means that I also excluded the inflected infinitive, which is morphologically marked in Portuguese, but was infrequent in my dataset.
- 2) I eliminated high frequency expressions made with a verb, such as *quer dizer* ‘I mean’, *sei lá* ‘I don’t know’, *não sei quê* ‘or whatever’, *tá(s) a ver* ‘you see’,

<sup>11</sup> It is the first and third persons singular that present similar verb forms in certain verb tenses, as for instance the verb *comer* ‘to eat’ in imperfect tense: *comia*, *comias*, *comia*, *comíamos*, *comiam*. In the past tense, this ambiguity between the first and third persons does not exist: *comi*, *comeste*, *comeu*, *comemos*, *comeram*.

<sup>12</sup> More precisely, the data come from web pages from four Portuguese-speaking countries: Angola, Brazil, Mozambique, and Portugal, and was collected in 2013–2014.

**Table 7:** Independent linguistic factors for coding SPE.

<b>Factors</b>	<b>Levels</b>
type of clause	main clause conjoined clause subordinated clause
priming effects	previous clause had a null subject previous clause had overt pronoun subject previous clause had an overt full NP subject no priming
morphological regularity	regular verb irregular verb
semantic content	mental activity verb stative verb external activity verb
verb form	complex verb form simple verb form
paradigm ambiguity	more obvious verb tense less obvious verb tense
person and number	first person singular second person singular third person singular first person plural third person plural
animacy	animate inanimate
coreferentiality	coreference with subject, no switch switch with subject, coreference with indirect object switch with subject, coreference with direct object switch with subject, coreference with oblique object complete switch
frequency (log10)	0–10 11–100 101–1 000 1 001–10 000 10 001–100 000 100 001–1 000 000 1 000 001–10 000 000 10 000 001–100 000 000

*digamos* ‘let’s say’, *como posso dizer* ‘how can I say’, and *sabe(s)* ‘you know’. These expressions are probably processed as one word, and not a sentence (cf. Heine, Claudi & Hünemeyer 1991; Heine & Kuteva 2003).

- 3) Following Otheguy and Zentella (2012), I excluded the verbs *ser* and *estar* ‘to be’ when they have no subject and mean ‘it’s’ (e.g. *é um bolo de chocolate* ‘it’s a chocolate cake’, *está bem* ‘it’s fine’) because they always take the third person singular agreement, but I included them when copulative (e.g. *é o meu irmão* ‘he’s my brother’, *você está longe* ‘you are far’) because these cases are marked for agreement between the subject and the verb.

- 4) I excluded the existential verbs *haver* and *ter* ‘to have’, because they have no subject. For instance, *há palavras que eu não percebo* ‘there are words that I don’t understand’, or *tem muita gente* ‘there is a lot of people’.
- 5) I excluded the verbs that have no subject (the ones that have expletive subjects in English, for example), such as *tá chovendo* ‘it’s raining’, or *neva* ‘it snows’.
- 6) I included incomplete clauses when the expression of the subject was clear. For example, I included *ele tem que...* ‘he has to...’ because the overt subject is clearly expressed, but I excluded *vivia...* ‘lived...’, because the SPP could be expressed after the verb, as in *vivia ele* ‘he lived’.
- 7) In Portuguese, when the antecedent of the relative clause is co-indexed with the subject of the relative clause, a null subject is expected, since subject resumption is uncommon. For instance, the token *tinha* in the sentence *batia numa filha que tinha cinco anos* ‘he beat a daughter who was five years old’ would not be included in the dataset because the antecedent of *tinha* is *filha* – they are co-indexed. However, the token *tinha* in the sentence *batia numa filha que ele tinha* ‘he beat a daughter that he had’ would be included because it is not co-indexed with *filha*, but rather with the subject of the verb *batia* ‘he beat’.

#### 4. Distribution of null and overt subject pronouns

One hundred tokens per speaker were coded (N = 5600). The overall distribution of subjects in my dataset was 55.2% null subjects, 25.4% overt subjects, and 19.4% full noun phrase subjects, as shown in **Table 8**.

To be consistent with other studies and because the focus of this section is pronominal use, I removed from the dataset the tokens with full noun phrase subjects, and retained only the subject pronouns, whether they are expressed or not (N = 4,512). To **Table 4**, I now add the results regarding the use of null and overt SPP in Santomean Portuguese (**Table 9**).

At first glance, the results suggest that the pronominal use in Santomean Portuguese is more similar to European Portuguese than to Brazilian Portuguese. In fact, 68.5% of subject pronouns are unexpressed in Santomean, which compares to 78% in European Portuguese and 44% in Brazilian Portuguese.<sup>13</sup> **Table 9** shows that the African varieties of Portuguese are situated between European and Brazilian Portuguese in their use of the null subject. However, remember that the methodologies and variables that underlie these studies are incomplete or unbalanced, and different from the ones used in the current study. Further studies with comparable data, methodology, and variables are necessary to validate this finding.

The following section deals with the social and linguistic constraints on the use of null subject.

**Table 8:** General distribution of subject expression (N = 5,600).

Coding	%
Null subject	55.2
Overt subject	25.4
Full noun phrase subject	19.4

<sup>13</sup> SPE might vary in Brazil from one region to another, from one study to another, but generally speaking, the rate of null subjects is always lower than in European Portuguese.



**Table 9:** SPE in European, Angolan, Santomean, Mozambican, and Brazilian Portuguese.

Variety	Null subjects	Overt subjects
EP	78.0%	22.0%
AP	75.1%	24.9%
STP	68.5%	31.5%
MP	52.5%	47.5%
BP	44.0%	56.0%

## 5. Results and analysis

The variation of SPE was modeled through logistic mixed-effects regression using the R package. The R package has the advantage of allowing random and mixed effects, which takes into account that some speakers might favor a linguistic outcome while others might disfavor it, regardless of what their social characteristics would predict, and that some words might be treated distinctively (Johnson 2009).<sup>14</sup> Rbrul was also used to perform one-level analyses and obtain factor weights, a statistical measure often used in sociolinguistics that indicates to what degree a variable is favored or disfavored.

The variation of SPE was modeled through a logistic mixed-effects with SPEAKER as a random effect using R package *lme4* (version 1.1-12; Bates et al. 2015). Both social and linguistic predictors of SPE, run as binary variables, were investigated. A backward elimination with the *anova* function was performed to find the best model. According to the best-fit model, the constraints EDUCATION LEVEL, TYPE OF CLAUSE, PRIMING EFFECTS, MORPHOLOGICAL REGULARITY, SEMANTIC CONTENT, PERSON AND NUMBER, ANIMACY, and COREFENTIALITY were all significant. The random SPEAKER effect is also significant ( $p < 0.001$ ). The factors AGE, GENDER, SPOKEN LANGUAGE, ETHNIC ORIGIN, VERB FORM, PARADIGM AMBIGUITY and FREQUENCY<sup>15</sup> were not significant. Each of the significant factors will be discussed one at a time.

### 5.1. Significant social factors for the use of null subject

In the full model with all factors included, EDUCATION and ETHNIC ORIGIN were almost but not quite significant, but when doing the backward elimination in R, deleting EDUCATION LEVEL would make ETHNIC ORIGIN significant, and eliminating ETHNIC ORIGIN would make EDUCATION LEVEL significant. This is usually a sign of correlation. The cross-tabulation in **Table 10** shows why there is interaction between EDUCATION LEVEL and ETHNIC ORIGIN.

The distribution of speakers is clearly skewed; those with just one Forro parent tend to be less well educated, so that most of the data for university educated speakers comes from those with two Forro parents, while data for speakers with only a primary education includes a high proportion of subjects with only one Forro parent.<sup>16</sup> Consequently, I decided to remove ETHNIC ORIGIN from the analysis and keep EDUCATION LEVEL. Without ETHNIC ORIGIN as a factor, EDUCATION LEVEL is significant.

<sup>14</sup>This means that if a word has a distinctive feature or characteristic, the R package will not attribute it more weight than any other words. To explain what this idiosyncratic treatment is, Johnson (2009: 381) gives the following example: “The most frequent loanword in the corpus is *dialekt*, which occurs 35 times. On the other hand, some 200 words only occur once each. If *dialekt* behaved more or less idiosyncratically, it would not make sense to weight it 35 times more heavily than any of the 200 or so words that only occur once. Therefore, a thorough mixed-model analysis of this data would include a random effect for word. This would lead to different estimates of speaker-internal factors such as orthography and word frequency.”

<sup>15</sup>Note that FREQUENCY was investigated as a continuous and discrete factor, with and without random speaker effect, and it never appeared as significant.

<sup>16</sup>Interestingly, this table shows that there is a Forro-elite that possibly resists ethnic mixing, so that mixing predominates in the lower-classes (bottom-up).



**Table 10:** Cross-tabulation between EDUCATION LEVEL and ETHNIC ORIGIN: number of tokens and percentage of data.

EDUCATION LEVEL	ETHNIC ORIGIN			
	two Forro parents		one Forro parent	
	N	%	N	%
primary school	914	63.6	524	36.4
high school	1623	87.4	233	12.6
university	1145	94.0	73	6.0

**Table 11:** Significant social factors for the use of null subject (intercept = 1.48; N = 4512; [ $\emptyset$ ] = 68.5%).

	Estimate	p-value	Factor weight	%[ $\emptyset$ ]	N-Total
<b>Education level</b>					
(vs. primary school)			0.55	73.9	1438
high school	-0.23	0.20	0.50	67.8	1856
university	-0.40	0.04*	0.45	63.2	1218
			range: 0.10		

Little social significance seems to be attached to SPE. In fact, participants did not address this feature during interviews (contrary to other features, such as pronunciation of rhotics which is more often mentioned; Bouchard 2017). As the results show, the conditioning effects of social predictors on SPE do not appear to be as strong as the linguistic predictors. The only social factor that appears as significant is EDUCATION LEVEL. And surprisingly, the results are contrary to expectations and previous studies (e.g. Ávila-Jiménez 1996): having a university degree disfavors the use of null subjects (factor weight: 0.55), while having no more than a primary school education favors it (factor weight: 0.45) (Table 11).

Speakers generally associate formality with standard language. My supposition is that overt subject in São Tomé is somehow associated with formality, or to “more proper” speech. Although SPE was never discussed during interviews or in informal conversations, grammatical ideology might be an explanation for the higher rate of overt subjects among highly educated Santomeans. Kroch and Small argue that this grammatical ideology prescriptively “favors the most direct correspondence between propositional form and surface syntax” (1978: 48); overt subject pronouns provide an explicit surface realization of a propositional form. Consequently, the results suggest that people with a higher level of education show greater adherence to the grammatical ideology of a standard language that favors the use of overt pronouns in Santomean Portuguese.

However, this finding arises many questions. First, Kroch and Small (1978) suggest that people do have prescriptive grammatical intuitions and that consciousness of the prestige norms can influence speech. In São Tomé and Príncipe, the prestige variety is still considered to be European Portuguese, which favors the use of null subjects. Consequently, if prescriptive grammatical intuitions were an explanation for SPE in Santomean Portuguese, then one would expect educated Santomeans to have a greater use of null subjects, with a rate similar to native speakers of European Portuguese.

Second, since the local creoles lack null subject, one might expect speakers of a Santomean creole to show a greater use of overt subjects. In fact, among the adult participants, there is a high number of Santomeans with a low level of education who are bilinguals (i.e. people who learned creole when they were children and who still use it today).

**Table 12:** Cross tabulation between SPOKEN LANGUAGES and EDUCATION LEVEL: number of participants.

SPOKEN LANGUAGES	EDUCATION LEVEL		TOTAL
	primary school	university	
monolingual (Portuguese L1)	4	4	8
Portuguese L1, some creole L2	5	9	14
bilingual (Portuguese and creole)	7	3	10
TOTAL	16	16	32

However, it is erroneous to assume that highly educated Santomeans do not speak creole. In fact, among the highly educated adult participants, 19% are bilingual and 56% have knowledge of creole as a L2 (**Table 12**).

Third, could this greater use of overt subjects be related to the interview setting? The speech data included in this study was elicited in individual interviews, and these interviews were collected after I had spent a period of time (starting during the third month, more precisely) in São Tomé to ensure that the questions asked were relevant. The first plan was to structure the interviews in modules that included demographic questions, as well as questions related to family, childhood, schooling, social network, identity, and language attitude. But in reality, the recording sessions followed no predetermined structure. The scope of the conversations was not limited to the question models; participants elaborated on topics that interested them. Therefore, the interview setting or any relation between the question-answer pair probably did not affect the speech of the participants more than it would have in other studies on SPE. Also, in order to mitigate the observer's paradox (Labov 1972b) and to collect casual speech, the tokens of SPP were taken from the middle of the interview.

Fourth, could this finding be related to the interviewer's variety of Portuguese? The interviewer (myself) speaks Brazilian Portuguese as a L2. Could these two elements of information (being a non-native speaker and speaking a Brazilian variety of Portuguese) have influenced the use of SPE of the participants? Were speakers with a higher level of education, i.e. speakers with a greater knowledge of the prescriptive grammar, adapting to the interviewer? If this were the case, the results regarding level of education and greater use of overt subjects would be due to the effects of external conditioning factors, and could not be considered a finding. However, because little social significance is attached to SPE, and because many of my participants did not know that I was not a native speaker of Portuguese, the Brazilian accent might have influenced the results more (if it did at all) than the fact that Portuguese is one of my L2, because I have a good command of the language. To my knowledge, there is no evidence of a register effect that endows educated speakers with the ability to use a different register with more null subjects in formal contexts. Further studies on the matter would be relevant, as they could clarify whether the correspondence between propositional form and surface syntax is favored in formal contexts.

## 5.2 Significant linguistic factors for the use of null subjects

A total of seven linguistic constraints significantly condition SPE in Santomean Portuguese, as the results presented in **Table 13** indicate. The significance of most of these factors is consistent with findings from previous studies, most specifically in the extensive literature on the topic in Spanish (cf. Cameron 1992; Orozco 2015a, 2015b, 2016; Otheguy & Zentella 2012; among others).

**Table 13:** Significant linguistic factors for the use of null subject (intercept = 1.48; N = 4512; [ $\emptyset$ ] = 68.5%).

	Estimate	p-value	Factor weight	%[ $\emptyset$ ]	N-total
<b>Type of clause</b>					
(vs. coordinate clause)			0.52	68.7	470
main clause	0.08	0.50	0.54	71.2	3170
subordinate clause	-0.35	<0.01**	0.44	58.4	872
			range: 0.10		
<b>Priming effects</b>					
(vs. no priming)			0.47	64.8	1009
full NP	0.17	0.17	0.52	63.1	604
null subject	0.43	<0.001***	0.58	76.9	1975
overt subject	-0.18	0.08	0.43	58.0	924
			range: 0.15		
<b>Morphological regularity</b>					
(vs. irregular)			0.54	71.1	1822
regular	-0.29	<0.001***	0.46	66.7	2690
			range: 0.08		
<b>Semantic content</b>					
(vs. external activity verb)			0.55	72.1	2773
mental activity verb	-0.08	0.46	0.53	64.3	603
stative verb	-0.49	<0.001***	0.43	61.8	1136
			range: 0.12		
<b>Person and number</b>					
(vs. 1st person singular)			0.46	64.5	1595
2nd person singular	-0.06	0.86	0.45	58.7	46
3rd person singular	0.31	<0.001***	0.54	73.5	2037
1st person plural	-0.02	0.86	0.46	54.7	406
3rd person plural	0.50	<0.001***	0.59	72.4	428
			range: 0.14		
<b>Animacy (vs animate)</b>					
inanimate	1.73	<0.001***	0.30	67.4	4309
			range: 0.40		
<b>Coreferentiality</b>					
(vs coreference with subject)			0.72	77.8	2491
Switch, coreference with IO	-2.04	<0.001***	0.25	33.3	21
Switch, coreference with DO	-0.82	0.01*	0.54	61.7	47
Switch, coreference with OO	-0.90	0.05*	0.52	60.9	23
Complete switch	-1.03	<0.001***	0.48	57.1	1930
			range: 0.47		

**TYPE OF CLAUSE.** The effect of TYPE OF CLAUSE on SPE is weak, but significant. Coordinate and main clauses favor the use of null subject (respective factor weights: 0.52 and 0.54), while subordinate clauses disfavor it (factor weight: 0.44). This suggests that coordinate and main clauses can be grouped together, as they are not significantly different. The results of this new recoding is presented in **Table 14**.

**Table 14:** The significance of the use of null subject for TYPE OF CLAUSE recoded (coordinate and main clauses vs. subordinate clauses) (intercept = 1.54; N = 4512;  $[\emptyset]$  = 68.5%).

	Estimate	p-value	Factor weight	% $[\emptyset]$	N-total
<b>Type of clause</b>					
(vs. coordinate and main)			0.55	70.9	3640
subordinate	-0.43	<0.001***	0.45	58.4	872
			range: 0.10		

The following sentences from my dataset illustrate these tendencies in the use of the pronouns:

- (10) Coordinate clause slightly favors the use of null subject  
 STP: Conheço Angolar **mas amigos Angolar**  $[\emptyset]$  **não tenho.**  
 ENG: know.1SG Angolar but friends Angolar NEG have.1SG  
 ‘I know Angolares, but I have no Angolar friends.’  
 - Yuri, 30 years old

- (11) Main clause slightly favors the use of null subject  
 STP: Esse rapaz lá  $[\emptyset]$  **estou habituada com ele já.**  
 ENG: this boy/guy there am used with him already  
 ‘This guy, I’m used to him already.’  
 - Mily, 20 years old

- (12) Subordinate clause disfavors the use of null subject  
 STP: É uma língua **que eu gostaria de aprender.**  
 ENG: it.is a language that I would.like.1SG to learn  
 ‘It’s a language I would like to learn.’  
 - Fábio, 26 years old

These results are similar to Orozco’s (2015a), with coordinate clauses favoring null subject and subordinate clauses disfavoring it. Otheguy and Zentella (2012), and Shin and Montes Alcalá (2014) also found that the null subject is favored in coordinate clauses.

PRIMING EFFECTS. Results regarding PRIMING EFFECTS show that the null subject is favored when the subject of the preceding clause is also null (factor weight: 0.58). It is overt subjects in previous clauses that disfavor null subjects the most (factor weight: 0.43). The following two sentences represent this pattern in Santomean Portuguese:

- (13) Null subject attracts null subject  
 STP:  $[\emptyset]$  quero falar com uma pessoa que  $[\emptyset]$  acho que  $[\emptyset]$   
 ENG: want.1SG to.talk with a person that think.1SG that  
 percebe pouco  
 understand.3SG little  
 ‘I want to talk to someone who I think doesn’t understand much’  
 - Anita, 69 years old

- (14) Overt subject attracts overt subject  
 STP: **eles** precisavam de viver, então o que que **eles** faziam...  
 ENG: they needed.3PL to live, so what that that they did.3PL  
 ‘they needed to live, so what would they do...’  
 - Tomás, 50 years old

**Table 15:** The significance of the use of null subject for PERSON AND NUMBER recoded (3rd persons vs. others) (intercept = 1.54; N = 4512; [ $\emptyset$ ] = 68.5%).

	Estimate	p-value	Factor weight	%[ $\emptyset$ ]	N-total
<b>Person and number</b>					
(vs. 1st and 2nd persons)			0.46	62.4	2047
3rd persons	0.36	<0.001***	0.54	73.5	2465
			range: 0.08		

This pattern agrees with Orozco (2015a: 204) who wrote about Barranquilla and New York Spanish that “one specific type of subject promotes the occurrence of subjects of the same type with overt pronominal subjects promoting overt subjects and null subjects promoting null subjects.”

MORPHOLOGICAL REGULARITY. In Santomean Portuguese, verb forms with irregular morphology favor the use of null subject (factor weight: 0.54), while verb forms with regular morphology disfavor it (factor weight: 0.46).

- (15) Irregular verbs favor the use of null subject

STP: tenho que lavar todos os outros dias

Eng: have.1SG to wash all the other days

‘I have to wash (the dishes) all the other days’

- Natália, 33 years old

- (16) Regular verbs disfavor the use of null subject

STP: nós herdamos muito dos Portugueses

Eng: we inherited.1PL a.lot from.the.PL Portuguese.PL

‘we inherited a lot from the Portuguese’

- Catarina, 43 years old

This finding, which is consistent with previous studies (Erker & Guy 2012), is probably related to the fact that irregular verbs often have distinctive forms for its different persons and numbers (e.g. the verb *ser* ‘to be’: *sou, és, é, somos, são*).

SEMANTIC CONTENT. External and mental activity verbs favor the use of null subjects with respective factor weights of 0.55 and 0.53, while stative verbs disfavor its use, with a factor weight of 0.43. One explanation to the disfavoring of null subject with stative verbs might be the high frequency of at least two stative verbs: *ser* and *estar* ‘to be’. Those two verbs show morphological irregularity in most of their inflectional forms, and as seen above, irregular forms of verb favor the use of null subject.

- (17) External activity verb favors the use of null subject

STP: trabalhava na padaria

Eng: used.to.work.3SG at.the bakery

‘he used to work at the bakery’

- Max, 24 years old

- (18) Mental activity verb favors the use of null subject

STP: é, acho que é isso que dificulta

Eng: yeah, think.1SG that it.is that that complicate.3SG

‘yeah, I think that’s what complicates things’

- Pilar, 44 years old

(19) Stative verb disfavors the use of null subject

STP: **eu** era casada  
 Eng: I was.1SG married  
 'I was married'

- Sandra, 38 years old

These results differ from Erker and Guy’s (2012) findings about Dominican Spanish in New York, but they are in agreement with Orozco’s (2015a) who found that external activity verbs favor the use of null subjects in Barranquilla and New York Spanish. However, Orozco (to appear) delved into the relation between semantic content and frequency of verbs in Caribbean Colombia Spanish spoken in New York City, and found that not all verbs within each semantic category behave the same. The general tendencies regarding SEMANTIC CONTENT and use of SPE (as shown in examples 17 to 19) might be skewed by the presence of some high frequency verbs. For instance, in Orozco’s study, the mental activity verbs *pensar* ‘to think’ and *creer* ‘to believe’ favor the use of null subject with respective factor weights of 0.78 and 0.64, and they represent 2.2% and 3.9% of all tokens in his dataset. As he wrote, these verbs are catalysts of the favorable effect on null subject use.

PERSON AND NUMBER. Null subjects are favored in third person singular and plural, with factor weights of 0.54 and 0.59 respectively, and are disfavored in all other positions, with factor weights of 0.45 and 0.46. These results suggest that there are two different groups with different tendencies: the third persons favor null subject, and the first and second persons disfavor it. A recoding of these levels underlines these tendencies (**Table 15**).

These results are contrary to expectation since the third person singular is morphologically unmarked, and therefore the overt subject does not appear as redundant. However, this is consistent with Barbosa, Duarte and Kato (2005) who have shown that the decrease of null subjects in Brazilian Portuguese has affected the first person (82% overt subject) and second person (78% overt subject) more than the third person (45% overt subject). These authors noted that “[t]his distinct behavior of the third person null subject led some Brazilian linguists to consider it a different type of empty category. Thus, for Figueiredo Silva (1996), Negrão & Müller (1996) and Modesto (2000b) it is a variable, and for Ferreira (2000) and Rodrigues (2004) it is a trace of A-movement” (2005: 46). That being said, results in Duarte (2000) show that both of the varieties of European and Brazilian Portuguese that she investigated present a lower rate of overt subjects in third person singular, i.e. a higher rate of null subject in this position. Remember that Duarte (1993) demonstrated an increasing rate of overt subjects over time in Brazilian Portuguese, and suggested that SPE might be undergoing change. This finding regarding PERSON AND NUMBER could be related to animacy and to the fact that third person pronouns are previously anchored in discourse (Duarte 1995, 2000). However, as seen in **Table 16**, null inanimate subjects in the first and second persons are frequent, and so, in a slightly higher percentage than in the third persons.

ANIMACY. This constraint significantly and strongly conditions SPE in Santomean Portuguese. The use of a null subject is favored when the subject of the verb is inanimate

**Table 16:** Cross tabulation between ANIMACY and PERSON AND NUMBER for the use of null subject: number of participants and percentages.

	1st persons		2nd person		3rd persons		TOTAL	
	N	%	N	%	N	%	N	%
animate	1878	77	4	22	1023	55	2905	67
inanimate	74	96	3	100	108	88	185	91
TOTAL	1952	78	7	33	1131	57	3090	68



(factor weight: 0.70), and disfavored when the subject of the verb is animate (factor weight: 0.30). These tendencies are illustrated in the following examples from my interviews:

- (20) Use of null subject with inanimate referent  
 STP: já não é forte, já não é forte  
 ENG: now NEG is strong, now NEG is strong  
 ‘it’s not strong anymore, it’s not strong anymore’  
 - Pilar, 44 years old

- (21) Use of overt subject with animate referent  
 STP: eu percebo dialeto  
 ENG: I understand.1SG dialect  
 ‘I understand Forro’  
 - Elzo, 50 years old

These results are consistent with the findings of Barbosa, Duarte and Karo (2005: 23) for European Portuguese: “One major condition that contributes to the difference between null and overt pronouns is animacy. In this regard, the results are striking. When the referent is [-animate], [European Portuguese] shows, in the sample analyzed, 97% of null subjects.” This is comparable to my sample of Santomean Portuguese in which 91% of the inanimate referents are expressed with null subjects.

COREFERENTIALITY. As expected, a null subject is favored (factor weight: 0.72) when there is complete coreference with the subject of the previous clause, and it is disfavored when there is a switch of reference (factor weight: 0.48). Interestingly, coreference with objects behaves differently according to the type of object: coreference with a direct object or an oblique object slightly favors the use of a null subject (respective factor weights of 0.54 and 0.52), but coreference with an indirect object strongly disfavors it (factor weight: 0.25). The following are illustrations of these patterns:

- (22) Coreference favors null subject  
 STP: fiz todo estágio, e depois quando [Ø] regresssei [Ø]  
 ENG: did.1SG all internship and after when came.back.1SG  
 fiquei no ISP  
 stayed.1SG in.the ISP  
 ‘I did the entire internship, and then when I came back I stayed at the ISP’<sup>17</sup>  
 - André, 46 years old

- (23) Coreference with an indirect object disfavors a null subject  
 STP: você vai dar lugar ao teu irmão porque ele vai busc...  
 ENG: you will.3SG give place to your brother because he will.3SG get...  
 ele vai vir para tomar lugar  
 he will.3SG come to get place  
 ‘you will give your place to your brother because he will get... he will come and occupy the place’  
 - Michel, 22 years old

- (24) Coreference with a direct object favors a null subject  
 STP: ela odiava-me, [Ø] odiava ela  
 ENG: she hated.3SG me, hated.1SG her  
 ‘she hated me, I hated her’  
 - Maria, 31 years old

<sup>17</sup> ISP is the *Instituto Superior Politécnico*, now called the *Universidade de São Tomé e Príncipe*.

- (25) Coreference with an oblique object favors a null subject  
 STP: **eu** brincava com meus primos, até [Ø] foram embora há muito tempo  
 ENG: I played.1SG with my cousins, even left.3PLthere.is a.lot.of time  
 ‘I used to play with my cousins, although they left a long time ago’  
 - Marcela, 12 years old
- (26) Complete switch in reference disfavors a null subject  
 STP: o que **eles** falam **eu** não...não entendo mesmo  
 ENG: what that they speak.3PL I NEG...NEG understand.1SG at.all  
 ‘what they speak I don’t...I don’t understand at all’  
 - Flor, 43 years old

However, as seen in **Table 13**, there are very few tokens of coreference with a complement (total of 91 tokens, representing 2% of all tokens). To get a clearer picture of this factor, I collapsed the levels of COREFERENTIALITY to make a binary distinction between no switch in reference, and switch in reference, with this last category including the complete switch and the partial switch with coreference with complements. This coding follows Erker and Guy (2012) (**Table 17**).

Results in **Table 17** show clearly that the use of a null subject is favored when there is no switch in reference (factor weight: 0.62, and 20.8% more null subjects).

Finally, **Table 18** is an updated version of **Table 13**; it presents the significant linguistic factors for SPE with the revised and combined factor groups as discussed throughout this chapter. This table gives a definitive and clearest picture of the meaningful linguistic constraints on the process.

## 6. Discussion and conclusion

Santomean Portuguese has a high rate of null subjects (68.5%), which makes it more similar to European Portuguese (78%) than to Brazilian Portuguese (44%) in its use of pronouns. Interestingly, when comparing different varieties of Portuguese, we see that the African varieties are situated between the European and Brazilian ones (as seen in **Table 9**). As is the case for Spanish, this suggests that SPE can serve as a tool to differentiate Portuguese varieties.

I examined the effects of five social and ten linguistic constraints. Results show that EDUCATION LEVEL, TYPE OF CLAUSE, PRIMING EFFECTS, MORPHOLOGICAL REGULARITY, SEMANTIC CONTENT, PERSON AND NUMBER, ANIMACY, and COREFERENTIALITY significantly condition SPE in Santomean Portuguese. The random SPEAKER effects were also significant.

Regarding social constraints, Santomeans with a lower level of education favor the use of null subject, and the ones with a higher level of education disfavor it. I suggest that the use of overt subject gives a direct correspondence between surface syntax and propositional form, which might explain this preference among highly educated Santomeans. Following

**Table 17:** The significance of the use of null subject for COREFERENTIALITY recoded (switch in reference vs. no switch in reference) (intercept = 1.54; N = 4512; [Ø] = 68.5%).

	Estimate	p-value	Factor weight	%[Ø]	N-total
<b>Coreferentiality</b>					
(vs no switch in reference)			0.62	77.8	2491
switch in reference	-1.04	<0.001***	0.38	57.0	2021
			range: 0.26		

**Table 18:** Significant linguistic factors for the use of null subject with recoding (intercept = 1.54; N = 4512; [ $\emptyset$ ] = 68.5%).

	Estimate	p-value	Factor weight	%[ $\emptyset$ ]	N-total
<b>Type of clause</b>					
(vs. main and coordinate clause)			0.55	70.9	3640
subordinate clause	-0.43	<0.001***	0.45	58.4	872
			range: 0.10		
<b>Priming effects</b>					
(vs. no priming)			0.48	64.8	1009
full NP	0.17	0.17	0.51	63.1	604
null subject	0.42	<0.001***	0.58	76.9	1975
overt subject	-0.20	0.06	0.43	58.0	924
			range: 0.15		
<b>Morphological regularity</b>					
(vs. irregular)			0.54	71.1	1822
regular	-0.28	<0.001***	0.47	66.7	2690
			range: 0.07		
<b>Semantic content</b>					
(vs. external activity verb)			0.55	72.1	2773
mental activity verb	-0.08	0.45	0.53	64.3	603
stative activity verb	-0.49	<0.001***	0.43	61.8	1136
			range: 0.12		
<b>Person and number</b>					
(vs. 1st and 2nd persons)			0.46	64.5	2047
3rd persons	0.36	<0.001***	0.54	58.7	2465
			range: 0.08		
<b>Animacy (vs animate)</b>					
inanimate	1.68	<0.001***	0.69	91.1	203
			range: 0.38		
<b>Coreferentiality</b>					
(vs no switch in reference)			0.62	77.8	2491
switch in reference	-1.04	<0.001***	0.38	57.0	2011
			range: 0.26		

Kroch and Small (1978) and their research on grammatical ideology, overt subject could also be associated with formality and standard language in Santomean Portuguese. Hence people with a higher level of education associate the use of overt subjects with formality and proper speech. An interview setting is certainly a context that favors the use of the standard, although the objective was to have access to the vernacular. However, this would be surprising as European Portuguese (which shows preference for null subject) is still considered to be the prestigious norm in São Tomé and Príncipe.

The conditioning effects of linguistic predictors on SPE appear to be stronger than the social predictors. Regarding TYPE OF CLAUSE, the use of a null subject is slightly favored in coordinate and main clauses. PRIMING EFFECTS also constrain the use of SPE; the results show that the use of one specific type of subject in one clause “attracts” the use of the same type of subject in the following clause. MORPHOLOGICAL REGULARITY also conditions the use of the pronouns, with irregular verbs favoring the use of a null subject. This is not

surprising as overt subjects with irregular verb forms may appear as being redundant. Irregular verbs give a more salient information independently of the use of the pronoun. As to SEMANTIC CONTENT of verb, external and mental activity verbs favor the use of a null subject. However, evidence from Orozco (to appear) suggests that the high frequency of certain verbs within each semantic category influences these results. If he is right, then the favoring of null or overt subjects would be a characteristic of each verb, and not of any semantic category. For PERSON AND NUMBER, the third persons behave quite differently than the other ones. A recoding of the third persons versus the others shows that they favor the use of a null subject. This supports the studies regarding Brazilian Portuguese in which third person subjects also behaved differently, and which led linguists to see third person subjects as a different type of empty category (cf. Barbosa, Duarte and Kato 2005). ANIMACY strongly conditions SPE, with most of the inanimate objects expressed with a null subject. Finally, as was expected regarding COREFERENTIALITY, no switch in reference favors the use of a null subject, as the referent remains the same.

I conclude with the following question: what happens to the variety of Portuguese spoken in São Tomé and Príncipe now that it is an independent nation? Santomean Portuguese can be seen as a language in contact with Forro and other creole languages of the islands, but the use of these creoles is restricted. Today, the great majority of young Santomeans are monolingual Portuguese speakers. Most of the Portuguese colonists left the islands as soon as São Tomé and Príncipe became independent, just as they abandoned the rest of the collapsing Portuguese colonies in Africa in 1974–1975. Since then, Santomeans have had a greater access to education and means of communication in Portuguese (e.g. television, Internet), and greater social mobility (in part related to Santomean immigration to Portugal).

Within this emergence of the Santomean variety of Portuguese, some features show the influence of the creole (Afonso 2008; D'Apresentação 2013; Figueiredo 2010; Gonçalves 2010, 2016; Lima 2009), others such as the use of rhotics show innovation (Bouchard 2017), and others such as SPE show conservatism. This conservatism from European Portuguese probably emerged after the colonial period, with better education opportunities for Santomeans. Little (or no) social significance is attached to this feature. One reason for this might be the fact that it is a feature that maintained a similar use to European Portuguese. Explanations to most findings regarding SPE in Santomean Portuguese can be found in previous studies. The one element that diverges from previous studies in other Portuguese-speaking countries is the fact that in São Tomé, highly educated people use overt subjects more frequently than less educated people, while null subjects are highly favored in European Portuguese regardless of level of education.

### Competing Interests

The author has no competing interests to declare.

### References

- Ambulate, J.** (2008). *A aquisição de sujeitos obrigatórios numa língua de sujeito nulo* [Acquisition of mandatory subjects in a null subject language.]. (Unpublished master's thesis). Universidade Nova de Lisboa, Portugal.
- Afonso, B.** (2008). *A problemática do bilinguismo e ensino da língua portuguesa em São Tomé e Príncipe* [The problem of bilingualism and teaching of Portuguese in São Tomé and Príncipe.]. (Unpublished master's thesis). Universidade Nova de Lisboa, Portugal.
- Ávila-Jiménez, B.** (1996). *Subject pronoun expression in Puerto Rican Spanish: A sociolinguistic, morphological, and discourse analysis*. (Unpublished doctoral dissertation). Cornell University, USA.

- Bailey, G.** (2004). Real and apparent time. In P. Trudgill, J. K. Chambers, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 312–332). Oxford: Blackwell Publishing. DOI: <https://doi.org/10.1002/9780470756591.ch12>
- Bailey, G., Wikle, T., Tillery, J., & Sand, L.** (1991). The apparent time construct. *Language Variation and Change*, 3, 241–264. DOI: <https://doi.org/10.1017/S0954394500000569>
- Barbosa, P.** (2000). Clitics: A window into the null subject property. In J. Costa (Org.), *Essays in Portuguese Comparative Syntax* (pp. 31–93). New York: Oxford Press.
- Barbosa, P.** (2009). Two kinds of subject pro. *Studia Linguistica*, 63(1), 2–58. DOI: <https://doi.org/10.1111/j.1467-9582.2008.01153.x>
- Barbosa, P.** (2010). Overt subjects in raising and control complements and the null subject parameter. *Linguistic Society of America Annual Meeting*, Retrieved January 16, 2018, from: <https://journals.linguisticsociety.org/proceedings/index.php/ExtendedAbs/article/view/512>. DOI: <https://doi.org/10.3765/exabs.v0i0.512>
- Barbosa, P.** (2013). Partial pro-drop as null NP anaphora. In Y. Fainleib, N. Lacara, & Y. Park (Eds.), *Proceedings of the 41st Annual Meeting of the North East Linguistic Society*, 2, 71–85.
- Barbosa, P., Duarte, M. E., & Aizawa Kato, M.** (2001). A distribuição do sujeito nulo no português europeu e no português brasileiro [The distribution of null subject in European and Brazilian Portuguese.]. *Actas do XVI Encontro Nacional da Associação Portuguesa de Linguística*, 539–550. Lisbon, Portugal.
- Barbosa, P., Duarte, M. E., & Aizawa Kato, M.** (2005). Null subjects in European and Brazilian Portuguese. *Journal of Portuguese Linguistics*, 4, 11–52. DOI: <https://doi.org/10.5334/jpl.158>
- Bates, D., Maechler, M., Bolker, B., & Walker, S.** (2015). Fitting linear mixed effects models using *lme4*. *Journal of Statistical Software*, 67, 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Becker, K.** (2013). The sociolinguistic interview. In C. Mallinson, B. Childs, & G. van Herk (Eds.), *Data collection in Sociolinguistics* (pp. 91–100). New York: Routledge.
- Bouchard, M.** (2017). *Linguistic variation and change in the Portuguese of São Tomé*. (Unpublished doctoral dissertation). New York University, USA.
- Cameron, R.** (1992). *Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation*. (Unpublished doctoral dissertation). University of Pennsylvania, USA.
- Cameron, R., & Flores-Ferrán, N.** (2004). Perseverance of subject expression across regional dialects. *Spanish in Context*, 1, 41–65. DOI: <https://doi.org/10.1075/sic.1.1.05cam>
- Chambers, J., & Trudgill, P.** (1998). *Dialectology* (2nd edition). Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511805103>
- Chavagne, J.-P.** (2005). *La langue portugaise d'Angola. Études des écarts par rapport à la norme européenne du portugais* [Angolan Portuguese. A study of divergences from the Standard European Portuguese.]. (Unpublished doctoral dissertation). Université de Lyon 2, France.
- Cheshire, J.** (2004). Sex and gender in variationist research. In P. Trudgill, J. K. Chambers, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 423–443). Oxford: Blackwell Publishing. DOI: <https://doi.org/10.1002/9780470756591.ch17>
- Chomsky, N.** (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- CIA World Factbook.** (2017). *São Tomé and Príncipe*. Retrieved November 13, 2017, from: <https://www.cia.gov/library/publications/the-world-factbook/geos/tp.html>.
- Costa, J., Lobo, M., & Silva, C.** (2009). Null objects and early pragmatics in the acquisition of European Portuguese. *Probus*, 21, 143–162. DOI: <https://doi.org/10.1515/prbs.2009.005>



- D'Apresentação, P.** (2013). *Influência do crioulo Forro na escrita do Português na crianças em idade escolar em São Tomé* [The influence of Forro on the writing of Portuguese among school-aged children in São Tomé.]. (Unpublished master's thesis). Universidade de Bragança, Portugal.
- Davies, M., & Ferreira, M.** (2017). *Corpus do Português* [Portuguese Corpus.]. Retrieved November 3, 2017, from: <http://www.corpusdportugues.org/>.
- Dias, H.** (2009). O parâmetro do sujeito nulo em aprendentes do português em Moçambique [The Null Subject Parameter in learners of Portuguese in Mozambique.]. In H. Dias (Ed.), *Português moçambicano: Estudos e reflexões* (pp. 17–46). Maputo: Imprensa Universitária.
- Duarte, M. E.** (1993). Do pronome nulo ao pronome pleno: A trajetória do sujeito no português do Brasil [From null pronoun to overt pronoun: The trajectory of the subject in Brazilian Portuguese.]. In I. Roberts, & M. Aizawa Kato, *Português Brasileiro: Uma viagem diacrônica (Homenagem a Fernando Tarallo)* (pp. 107–128). Campinas: Editora da UNICAMP.
- Duarte, M. E.** (1995). *A Perda do Princípio “Evite pronome” no Português Brasileiro* [The loss of the principle “Avoid Pronoun” in Brazilian Portuguese.]. (Unpublished doctoral dissertation). UNICAMP, Brazil.
- Duarte, M. E.** (2000). The loss of the ‘Avoid Pronoun’ principle in Brazilian Portuguese. In M. Aizawa Kato & E. Negrão (Eds.), *Brazilian Portuguese and the null subject parameter* (pp. 17–36). Frankfurt am Main/Madrid: Vervuert/Iberoamericana.
- Eckert, P.** (1990). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, 1, 245–268. DOI: <https://doi.org/10.1017/S095439450000017X>
- Erker, D., & Guy, G.** (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language*, 88(3), 526–557. DOI: <https://doi.org/10.1353/lan.2012.0050>
- Ferreira, M.** (2000). *Argumentos Nulos em Português Brasileiro* [Null arguments in Brazilian Portuguese.]. (Unpublished master's thesis). UNICAMP, Brazil.
- Figueiredo, C. F. G.** (2010). *A concordância plural variável no sintagma nominal do português reestruturado da comunidade de Almoxarife, São Tomé* [Variable plural agreement in noun phrases in the restructured Portuguese of the community of Almoxarife, São Tomé.]. (Unpublished doctoral dissertation). University of Macau, China.
- Figueiredo Silva, M. C.** (1996). *A posição sujeito no português brasileiro: Frases finitas e infinitas* [The subject position in Brazilian Portuguese: Finite and non-finite clauses.]. Campinas: Editora da UNICAMP.
- Flores-Ferrán, N.** (2002). *A sociolinguistic perspective on the use of subject personal pronouns in Spanish narratives of Puerto Ricans in New York City*. Munich, Germany: Lincom-Europa.
- Fought, C.** (2004). Ethnicity. In P. Trudgill, J. K. Chambers, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 444–472). Oxford: Blackwell Publishing. DOI: <https://doi.org/10.1002/9780470756591.ch18>
- Galves, C.** (1990). Clíticos: Geometria ou álgebra, morfologia ou sintaxe. *DELTA*, 6, 255–271.
- Gonçalves, R.** (2010). *Propriedades de subcategorização verbal no português de São Tomé* [Properties of verbal subcategorization in Santomean Portuguese.]. (Unpublished master's thesis). Universidade de Lisboa, Portugal.
- Gonçalves, R.** (2016). *Construções ditransitivas no Português de São Tomé* [Ditransitive constructions in Santomean Portuguese.]. (Unpublished doctoral dissertation). Universidade de Lisboa, Portugal.



- Guy, G.** (1981). *Linguistic variation in Brazilian Portuguese*. (Unpublished doctoral dissertation). University of Pennsylvania, USA.
- Hagemeyer, T.** (in press). From creoles to Portuguese: Language shift in São Tomé and Príncipe. In L. Álvarez López, P. Gonçalves, & J. Avela (Eds.), *The Portuguese language continuum in Africa and Brazil*. Amsterdam/Philadelphia: John Benjamins.
- Heine, B., Claudi, U., & Hünnemeyer, F.** (1991). *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.
- Heine, B., & Kuteva, T.** (2003). On contact-induced grammaticalization. *Studies in Language*, 27(3), 529–572. DOI: <https://doi.org/10.1075/sl.27.3.04hei>
- Holmberg, A.** (2005). Is there a little *pro*? Evidence from Finnish. *Linguistic Inquiry*, 36(4), 533–564. DOI: <https://doi.org/10.1162/002438905774464322>
- Holmberg, A., Nayudu, A., & Sheehan, M.** (2009). Three partial null-subject languages: A comparison of Brazilian Portuguese, Finnish, and Marathi. *Studia Linguística*, 63(1), 59–97. DOI: <https://doi.org/10.1111/j.1467-9582.2008.01154.x>
- Instituto Nacional de Estatística (INE).** (2012). Resultados nacionais. *IV Resenseamento geral da população e da habitação* [National results. IV General population and housing census.]. Retrieved November 3, 2017, from: <http://www.ine.st/Documentacao/Recenseamentos/2012/DDENRAP/Resultados%20Nacionais%20do%20IV%20RGPH%202012%20novo.pdf>.
- Johnson, D. E.** (2009). Getting off the Goldvarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 3(1), 359–383. DOI: <https://doi.org/10.1111/j.1749-818X.2008.00108.x>
- Kato, M. A.** (1999). Strong pronouns and weak pronominals in the null subject parameter. *Probus*, 11(1), 1–37. DOI: <https://doi.org/10.1515/prbs.1999.11.1.1>
- Kroch, A., & Small, C.** (1978). Grammatical ideology and its effect on speech. In D. Sankoff (Ed.), *Linguistic variation: Models and Methods* (pp. 45–55). New York: Academic Press.
- Labov, W.** (1963). The social motivation of a sound change. *Word*, 19, 273–309. DOI: <https://doi.org/10.1080/00437956.1963.11659799>
- Labov, W.** (1966). *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- Labov, W.** (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W.** (1984). Field methods of the project on linguistic change and variation. In J. Baugh, & J. Sherzer (Eds.), *Language in Use: Readings in Sociolinguistics* (pp. 28–53). Englewood Cliffs, N.J.: Prentice-Hall.
- Lima, H. A.** (2009). *Interferências linguísticas: Um contributo para o ensino da língua portuguesa em São Tomé e Príncipe* [Linguistic interferences: A contribution to the teaching of Portuguese in São Tomé and Príncipe.]. (Unpublished master's thesis). Universidade de Lisboa, Portugal.
- Lobo, M.** (1995). *Para uma redefinição do parâmetro do sujeito nulo* [For a redefinition of the Null Subject Parameter.]. (Unpublished master's thesis). Universidade de Lisboa, Portugal. DOI: <https://doi.org/10.1515/9783110368840-023>
- Lobo, M.** (2016). Sujeitos nulos: Gramática do adulto, aquisição de L1 e variação dialetal [Null subjects: Adult grammar, L1 acquisition, and dialectal variation.]. In A. M. Martins, & E. Carrilho (Eds.), *Manual de linguística portuguesa* (pp. 558–580). Berlin/Boston: De Gruyter Mouton.
- Lucchesi, D., Baxter, A., & Ribeiro, I.** (2009). *O português afro-brasileiro* [Afro-Brazilian Portuguese.]. Salvador: EDUFBA. DOI: <https://doi.org/10.7476/9788523208752>

- Magalhães, T., & Santos, A. L.** (2006). As respostas verbais e a frequência de sujeito nulo na aquisição do Português Brasileiro e do Português Europeu [Verbal responses and null subject frequency in the acquisition of Brazilian and European Portuguese.]. *Letras de Hoje*, 41(1), 179–193.
- Martínez-Sanz, C.** (2011). *Null and overt subjects in a variable system: The case of Dominican Spanish*. (Unpublished doctoral dissertation). University of Ottawa, Canada.
- Modesto, M.** (2000a). *On the Identification of Null Arguments*. (Unpublished doctoral dissertation). University of Southern California, USA.
- Modesto, M.** (2000b). Null Subjects without “rich” agreement. In M. Aizawa Kato & E. Negrão (Eds.), *Brazilian Portuguese and the Null Subject Parameter*. Frankfurt: Vervuert Iberoamericana.
- Modesto, M.** (2009). Null subjects in Brazilian Portuguese: A critique of two possible analyses. In M. A. Torres Morais, & M. L. Andrade (Eds.), *História do português caipira* (pp. 99–123). Campinas: IEL Publicações.
- Morales, A.** (1997). La hipótesis funcional y la aparición de sujeto no nominal: El español de Puerto Rico. *Hispania*, 80(1), 53–65. DOI: <https://doi.org/10.2307/345995>
- Negrão, E.** (1997). Asymmetries in the distribution of overt and empty categories in Brazilian Portuguese. In J. R. Black & V. Motapanyase (Eds.), *Clitics, Pronouns and Movement: Current Issues in Linguistic Theory* (pp. 217–235). Amsterdam, Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/cilt.140.09neg>
- Negrão, E., & Müller, A. K.** (1996). As mudanças no sistema pronominal do português brasileiro: Substituição ou especialização [Changes in the pronominal system of Brazilian Portuguese: Substitution or specialization.]. *DELTA*, 12(1), 125–152.
- Oliveira, M. S. D., & Ferreira dos Santos, E.** (2007). Pronomes nulos na posição de sujeito no português de Angola: Um estudo preliminar [Null pronouns in subject position in Angolan Portuguese: A preliminary study.]. *Filologia e Linguística portuguesa*, 9, 85–102. DOI: <https://doi.org/10.11606/issn.2176-9419.v0i9p85-101>
- Orozco, R.** (2015a). *Spanish in Colombia and New York City: Dialectal parallelism meets language contact*. (Unpublished doctoral dissertation). New York University, USA.
- Orozco, R.** (2015b). Pronominal variation in Costeño Spanish. In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (Eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal Perspective* (pp. 17–37). Washington, DC: Georgetown University Press.
- Orozco, R.** (2016). Subject pronoun expression in Mexican Spanish: ¿Qué pasa en Xalapa? *Proceedings of Linguistic Society of America*. Retrieved November 3, 2017, from: <http://journals.linguisticsociety.org/proceedings/index.php/PLSA/article/view/3703>.
- Orozco, R.** (to appear). *El castellano del Caribe en la ciudad de Nueva York: Ele uso variable de sujetos pronominales* [Caribbean Castilian in New York City.].
- Otheguy, R., & Zentella, A. C.** (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199737406.001.0001>
- Otheguy, R., Zentella, A. C., & Livert, D.** (2007). Language contact in Spanish in New York toward the formation of a speech community. *Language*, 83(4), 770–802. DOI: <https://doi.org/10.1353/lan.2008.0019>
- Raposo, E.** (1986). On the Null Object Construction in European Portuguese. In O. Jaeggli, & C. Silva-Corvalán (Eds.), *Studies in Romance Linguistics* (pp. 373–390). Dordrecht: Foris.
- Rodrigues, C.** (2002). Morphology and null subjects in Brazilian Portuguese. In D. Lightfoot (Ed.), *Syntactic effects of morphological change* (pp. 160–178). Oxford: OUP. DOI: <https://doi.org/10.1093/acprof:oso/9780199250691.003.0009>

- Rodrigues, C.** (2004). *Impoverished morphology and A-movement out of case domains*. (Unpublished doctoral dissertation). University of Maryland, USA.
- Santos, E. F. dos.** (2006). *Reorganização do corpus de Chavagne (2005 – Anexo 1)* [Reorganization of Chavagne's corpus (2005 – Appendix 1)]. Manuscript.
- Seibert, G.** (2006). *Comrades, clients and cousins: Colonialism, socialism and democratization in São Tomé and Príncipe*. Leiden, Netherlands: Brill.
- Sheehan, M.** (2006). *The EPP and null subjects in Romance*. (Unpublished doctoral dissertation). Newcastle University, USA.
- Shin, N., & Alcalá, C. M.** (2014). El uso contextual del pronombre sujeto como factor predictivo de la influencia del inglés en el español en Nueva York [The contextual use of subject pronouns as a predictor of English influence on New York Spanish.]. *Sociolinguistic Studies*, 8, 85–110. DOI: <https://doi.org/10.1558/sols.v8i1.85>
- Spradley, J.** (1979). *The ethnographic interview*. New York: Holt, Rinehart, and Winston.
- Tagliamonte, S.** (2006). *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511801624>
- Teixeira, E. P.** (2013). A representação do sujeito pronominal no português popular angolano [The representation of pronoun subject in popular Angolan Portuguese.]. *PAPIA*, 22(1), 141–159.
- Travis, C.** (2005). The yo-yo effect: Priming in subject expression in Colombian Spanish. In E. J. Rubin, & R. Gess (Eds.), *Selected papers from the 34th Linguistic Symposium on Romance Languages (LSRL)* (pp. 329–349). Amsterdam, The Netherlands: John Benjamins.
- Trudgill, P.** (1974). *The social differentiation of English in Norway*. Cambridge: Cambridge University Press.
- Trudgill, P.** (2000). *Sociolinguistics: An introduction to language and society* (4th edition). London: Penguin Books.

**How to cite this article:** Bouchard, M.-E. (2018). Subject Pronoun Expression in Santomean Portuguese. *Journal of Portuguese Linguistics*, 17: 5, pp. 1–29, DOI: <https://doi.org/10.5334/jpl.191>

**Submitted:** 03 November 2017      **Accepted:** 10 April 2018      **Published:** 22 May 2018

**Copyright:** © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[ *Journal of Portuguese Linguistics* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 