

# THE PORTUGUESE LANGUAGE OF SÃO TOMÉ AND PRÍNCIPE<sup>1</sup>

A LÍNGUA PORTUGUESA DE SÃO TOMÉ E PRÍNCIPE

LA LENGUA PORTUGUESA DE SÃO TOMÉ Y PRÍNCIPE

**Amanda Macedo Balduino\***

University of São Paulo

**Manuele Bandeira\*\***

University for International Integration of the Afro-Brazilian Lusophony

**Shirley Freitas\*\*\***

University for International Integration of the Afro-Brazilian Lusophony

**ABSTRACT:** This paper provides a linguistic overview of the history and current status of Santomean and Principean Portuguese (PST and PP). The rise of these varieties in São Tomé and Príncipe dates back to their colonial origins in a multilingual environment, as well as to their development into Second Language of indigenous people — later nativized as mother language by their offspring. Besides, based on data collected in situ, this paper presents a brief synchronic overview of some phonological features of these varieties, focusing on phonological phenomena such as nasalization, rhotic variation, vocalization and segmental deletion, which are recurrent in PST and in PP. The analysis of historical, social, and linguistic factors allows us to conclude that the development of local varieties is evident in STP. Despite the lack of robust linguistic description and systematization, PST and PP are endowed with identity linguistic traits.

**KEYWORDS:** Portuguese. São Tomé and Príncipe. Phonological processes.

**RESUMO:** Este artigo discute a história e a situação linguística atual do português santomense e do português principense (PST e PP). O surgimento dessas variedades em São Tomé e Príncipe (STP) remonta ao estabelecimento colonial do português em um ambiente plurilíngue, bem como ao seu desenvolvimento inicial como segunda língua, e, posteriormente, como língua materna adquirida pelos nativos. Ademais, a partir de dados coletados *in loco*, o artigo apresenta um breve panorama sincrônico de alguns

---

<sup>1</sup> We thank FAPESP (grants no. 2015/25332-1 e 2017/26595-1) and CNPq (grant no. 150051/2018-2), whose funding enabled the development of this research and writing this paper.

\* Doctoral student in Philology and Portuguese – University of São Paulo. Email: [amanda.m\\_b@hotmail.com](mailto:amanda.m_b@hotmail.com).

\*\* Assistant Professor – University for International Integration of the Afro-Brazilian Lusophony (UNILAB), Malês campus. Professor – PPGLinC – Graduate Program in Language and Culture, Federal University of Bahia (UFBA). Email: [manuelebandeira@unilab.edu.br](mailto:manuelebandeira@unilab.edu.br).

\*\*\* Assistant Professor – University for International Integration of the Afro-Brazilian Lusophony (UNILAB), Malês campus. Professor – PPGLinC – Graduate Program in Language and Culture, Federal University of Bahia (UFBA). Email: [shirleyfreitas@unilab.edu.br](mailto:shirleyfreitas@unilab.edu.br).

processos fonológicos, a saber: nasalização, alternância do rótico, vocalização e apagamento segmental, observados no PST e no PP. A análise de fatores históricos, sociais e linguísticos nos permite concluir que, em STP, é evidente o desenvolvimento de variedades locais próprias, as quais, embora ainda careçam de uma descrição e sistematização linguística robusta, são dotadas de traços linguísticos identitários.

**PALAVRAS-CHAVE:** Língua portuguesa. São Tomé e Príncipe. Processos fonológicos.

**RESUMEN:** Este artículo analiza la historia y la situación lingüística actual del portugués de São Tomé y Príncipe (PST y PP). La aparición de estas variedades en São Tomé y Príncipe (STP) se remonta al establecimiento colonial del portugués en un entorno multilingüe, así como a su desarrollo inicial como segunda lengua y, posteriormente, como lengua materna adquirida por los nativos. Además, a partir de los datos recopilados in situ, el artículo presenta un breve panorama sincrónico de algunos procesos fonológicos, a saber: nasalización, alternancia de la rotica, vocalización y borrado segmentario, observado en el PST y PP. El análisis de factores históricos, sociales y lingüísticos permite concluir que, en STP, se evidencia el desarrollo de sus propias variedades locales, las cuales, si bien aún carecen de una descripción y sistematización lingüística robusta, están dotadas de rasgos lingüísticos identitarios.

**PALABRAS CLAVE:** Lengua portuguesa. São Tomé y Príncipe. Procesos fonológicos.

## 1 INTRODUCTION

This paper presents some characteristics of the Portuguese variety spoken in São Tomé and Príncipe (STP). We discuss the historical and social factors that led Portuguese to be acquired as majority and mother language (L1) in the country and the current political status of the Portuguese language in Santomean society. In addition, in face of the need for more studies on different Portuguese varieties spoken in Africa, we aim to present some synchronic phonological processes of Santomense Portuguese (PST) and Principense Portuguese (PP): Nasalization; rhotic alternation; vocalization; and segmental deletion.

As any other linguistic system, Portuguese is diverse, and its varieties showcase both shared and unique linguistic traits. This is the case of PST and PP, vernacular varieties used as mother language by the urban population of São Tomé and Príncipe. The purpose of this study is to discuss these varieties and to contribute to the registration and systematization of PST and PP, which is justified by the multilingual environment in which these varieties are inserted, the initial transmission from an input corresponding to a second language (L2), and the need for a linguistic description of PST and PP. Thus, we aim to analyze the confluence of historical, social, and linguistic factors, which does not only help understand the linguistic predominance of Portuguese in STP, but also allows the recognition of some particular and legitimate linguistic norms and patterns within the varieties that have emerged in the archipelago.

This paper is organized as follows: Section 2 lays out historical facts from the first and second periods of Portuguese colonization in STP (15<sup>th</sup>–16<sup>th</sup> century and 19<sup>th</sup>–20<sup>th</sup> century, respectively) and Section 3 brings a discussion on the post-colonial period, addressing the emergence of the Portuguese language of STP. Section 4 explains our methods of analysis and Section 5 discusses phonetic and phonological aspects that are inherent to the urban varieties of Portuguese spoken in the archipelago. Finally, Section 6 brings our final considerations.

## 2 PORTUGUESE PRESENCE IN SÃO TOMÉ AND PRÍNCIPE

São Tomé and Príncipe (STP) is a country in the Gulf of Guinea, on the West African coast, consisting of two islands — São Tomé and Príncipe. The Portuguese language is widely used by the population, characterizing STP as the only African country whose majority and mother language is Portuguese (GONÇALVES, 2010; GONÇALVES; HAGEMEIJER, 2015; BOUCHARD, 2017; BAXTER, 2018; ARAUJO, 2020; SANTIAGO; AGOSTINHO, 2020). Despite the massive use of Portuguese in STP, other languages also circulate in the islands, such as Santome (ISO 639-3: cri), Lung'le (ISO 639-3: pre), Angolar (ISO 639-3: aoa), and Kabuverdianu (ISO 639-3: kea) (FERRAZ, 1979; MAURER, 2009; HAGEMEIJER, 2009; 2011; ARAUJO; HAGEMEIJER, 2013; AGOSTINHO,

2015; BANDEIRA, 2017). The non-harmonious coexistence of Portuguese and the other languages is supported by historical and social factors, responsible for fostering a complex local linguistic ecology and structuring the archipelago's linguistic interactions.

The first Portuguese citizens settled on São Tomé and Príncipe Islands in 1493 (FERRAZ, 1979). The arrival of a larger contingent of settlers, all Portuguese speakers, led to the importation of enslaved labor for manual and domestic services. Enslaved newcomers were originally from different regions of the continent, such as the Niger Delta and the Kingdom of Congo, and spoke several African languages, possibly from the Kwa, Edo, Itsekiri, and Ijo groups (cf. HAGEMEIJER, 2009; BANDEIRA, 2017).

The interlinguistic contact in the archipelago, marked by the forced coexistence of different ethnicities, enabled the approximation of unintelligible African languages spoken by enslaved groups and the Portuguese of the 15<sup>th</sup> century, which, being the language of the colonizer — the dominant social group — was a possible linguistic target. Thus, Portuguese learning corresponded to one of the countless reflections of the strong assimilation policy employed by the Portuguese Crown in STP. The purpose of such a policy was to erase the enslaved population's ethnic identity by "Europeanizing" the colony and its inhabitants. Portuguese diffusion was one of the strategies used. In addition, such linguistic pressures were accompanied by other institutionalized racist actions, such as the mandatory baptism of the enslaved population since 1514 (SEIBERT, 2015). Within six months after settling in the archipelago, the enslaved group was given Christian, Europeanized names and was violently assimilated into Portuguese culture, having its ethnic and individual identities silenced.

After the settlement of enslaved groups and driven by assimilatory policies, D. Manuel determined in 1515 that each exiled man on the island should be granted an enslaved woman. Their offspring would be born free. This added complexity to the colony's social strata, as former slaves held a high status within the local society (BANDEIRA, 2017). The high social status of freed natives was seen as proof of success of the institutionalized colonial assimilation in STP. The Crown had forged an indigenous elite that had access to privileges denied to other groups and were willing to cooperate with the metropolis, even if marginalized by the Portuguese population. In addition, free slaves were also essential for the process of nativization of a pidgin spoken in the archipelago (HAGEMEIJER, 2009) and disseminated across the territory: The Proto-Creole of the Gulf of Guinea (PGG) (BANDEIRA, 2017).

According to Bandeira (2017), Fa d'ambô and Lung'le — spoken on Ano Bom Island and Príncipe Island, respectively — emerged from the PGG after the groups were taken to those two islands for colonial purposes. In their turn, Santome (or Forro), spoken in the capital, was a PGG daughter language that remained in the non-branched protolanguage's initial emergency site (BANDEIRA, 2017). Finally, fugitive enslaved people distanced themselves from the capital, organizing communities of maroon resistance that became the Angolan community — place of emergence of the Angolar.

Between the 17<sup>th</sup> and 18<sup>th</sup> centuries, the decline of the local economy led to a massive departure of Portuguese settlers from STP, which ended up favoring the use of Santome, Lung'le, and Angolar (FERRAZ, 1979; BANDEIRA, 2017). In this period, access to Portuguese was scarce and indigenous languages were widely spoken (cf. SEIBERT, 2012; ARAUJO, 2020).

The local crisis lasted until 1852, when Portugal started a second period of colonization in STP with the reinstallation of the plantation system, based on the culture of coffee and cocoa (cf. SEIBERT, 2015). This second colonization attempt was based on the employment of independent workers — STP received an expressive number of contractors, which drove a demographic change on the archipelago (SEIBERT, 2015). From the beginning of the 20<sup>th</sup> century until 1950, the population of contractors from Angola, Mozambique, and Cape Verde was more numerous than the native population of STP (NASCIMENTO, 2008). In relation to the white population, it was also possible to note a greater influx of colonists. Solely in the period between 1885–1900, the number of Portuguese citizens in the archipelago increased from 150 to 1,185 (SEIBERT, 2015). However, despite this increase in the Portuguese contingent, the white population accounted for only 2.8% of the total demographics of STP, which was negligible for Portuguese diffusion (ARAUJO, 2020).

Despite the large number of contractors, they do not account for the consolidation of the Portuguese language on the archipelago — on the contrary: The influx of contractors, especially from Cape Verde, intensified the multilingual scenario in STP. Were Cape Verdeans not only numerous, but they also settled in STP with their families and maintained Kabuverdianu as a vehicle and transmission language among this ethnic group, even though they were also subjected to assimilatory pressures, just like the local

population. Nascimento (2010), Seibert (2015), and Bouchard (2017) indicate that the influx of contractors intensified the discriminatory structure underlying social relations within the archipelago: While Portuguese citizens held political and economic power, segregationist issues caused by the assimilation policy were also evident among the local population.

Forros descended from an elite forged by assimilatory policies and, therefore, retained privileges denied to others. Nascimento (2010) and Bouchard (2017) state that, among the various groups that make up STP society, Forros had more prestige, being educated in Portuguese and occupying positions within public offices. In their turn, Principenses were seen in a friendly way, but as naive and easily deceived. Angolares received the labels of rude and savages. Finally, the contractors, who, like the Angolares, were concentrated in the fields and in interior regions, were seen as savages (SEIBERT, 2015).

Seibert (2015) points out that the discrimination suffered by the different groups was related to their level of assimilation of the Portuguese culture: The closer they were to Portuguese elements, the less discrimination they felt. Speaking Portuguese was one of the most valued elements. However, Portuguese proficiency was limited to a small local elite, as the general population had very limited access to the Portuguese language. Despite this scenario, assimilatory pressures promoted by the Portuguese Crown and the consolidation of social segments in STP led to a broader adoption of Portuguese in the islands and, consequently, to decreasing use and even threat of native languages, as discussed by Gonçalves (2010), Gonçalves and Hagemeyer (2015), Seibert (2015), Araujo (2020), and Santiago and Agostinho (2020). According to Araujo (2020), the diffusion of Portuguese in across STP cannot be explained only by the establishment of Portuguese as the country's official language after independence in 1975 (this being an important factor), but also by urbanization, schooling, and media diffusion.

The Urbanization Plan was an attempt to modernize and homogenize Portuguese overseas territories (cf. MILHEIRO, 2012). With that intent in mind, schools were brought to the city of São Tomé in 1951 as an urban apparatus (MILHEIRO, 2012). This was responsible for democratizing education and, since classes took place in Portuguese, disseminating the language across the country (GONÇALVES; HAGEMEIJER, 2015).

According to Nunes (2019), learning in a language and learning a language are distinct processes. In STP, the population with access to schools was educated in Portuguese and not in their mother languages. Exposure to indigenous languages was non-existent, since there were no linguistic policies concerned with their preservation at the time (ARAUJO, 2020). Soon, teaching in Portuguese — until then a foreign language — replaced teaching in native languages, a common process in places where European languages coexist with local languages. In this regard, Nunes (2019) states that European languages restrict and camouflage native languages. The mandatory nature of Portuguese and the erasure of other languages in STP schooling during the post-colonial period composed an exclusionary system that marginalized and silenced local languages, while access to schools was limited to those willing to speak Portuguese.

Another important aspect to understand the prevalence of Portuguese in STP concerns the participation of local elites (ARAUJO, 2020). According to Araujo (2020), the choice of Portuguese as official language was justified by a discourse of “ethnic neutrality” in view of the linguistic diversity of the archipelago. However, such choice obscured the fact that Portuguese was already the language of a social group, the Santomean elite. The characterization of linguistic diversity as a problem that can be solved with the establishment of Portuguese as official and vehicular language constitutes an argument that is still in force today in the PALOP (Portuguese-speaking African countries). Once seen as a natural phenomenon, linguistic diversity began to be seen as an anomalous and dysfunctional reality, both in ideology and in practice. Monteagudo (2012, p. 45) claims that this transformation is the result of “glotopoly operations to homogenize populations that speak several languages”. With the withdrawal of the Portuguese elite from power, the Santomean elite kept other groups of the archipelago — speakers of other autochthonous African languages or not — at the base of the social pyramid, preserving the values inherited from the assimilation policy in STP and, hence maintaining the status quo that once belonged to the Portuguese (SEIBERT, 2012; BOUCHARD, 2017; ARAUJO, 2020).

The success of Portuguese as L1 resulted from a confluence of factors. Urbanization, teaching democratization, and the absence of language policies in favor of indigenous languages forced the population to learn Portuguese and silenced local languages. The Santomean society was permeated by values inherited from the colonial policy of assimilation, initially disseminated by the

Lusitanian Crown and, in post-colonial period, propagated by the local elite. Thus, most native STP groups favored linguistic transmission of Portuguese, firstly acquired as L2 and then transmitted as linguistic input to younger generations, who in their turn acquired it as L1. Therefore, the speaker's linguistic attitude, shaped in the process of colonial socialization, was critical for the diffusion of Portuguese.

The adoption and use of a linguistic system are ideological acts (cf. NUNES, 2019). It is precisely the socially constructed ideologies that guide the linguistic use(s) of the speakers. In the case of STP, these ideologies are the ones that have fostered a negative attitude towards indigenous languages and a positive one towards Portuguese. Native groups distanced themselves from their linguistic identity in different ways.<sup>2</sup> The use of local languages reflected the individual's ethnic identity and, paradoxically, revealed belonging to a linguistic community of low status. Portuguese, on the other hand, presented as the language of national unification, was sold as "everyone's language", facilitating the alternation of linguistic code transmission. In other words: Despite the contemporaneity of the multilingualism phenomenon in STP, Portuguese outweighed local languages — legacy of a Lusophony culture once used for geopolitical control purposes and supported by assimilation policies.

In STP, Portuguese is the population's vehicular language of everyday use and, often, the speakers' only language, especially amongst the younger ones. The specificities of the multilingual conjuncture in STP have fostered the contact between local languages and Portuguese — when acquired as L2 and even today, when Portuguese is the majority L1 in the country. Besides, this scenario also culminated in the development of local varieties of Portuguese, with unique linguistic traits. The different Portuguese speeches of STP no longer correspond to the European standard initially established there, as they "*santomenizaram-se*" (got "Santomized"), corresponding nowadays to São Tomé and Príncipe's variety(ies) of Portuguese.<sup>3</sup> That said, the following Section gives a brief socio-historical overview of Portuguese as a particular variety of STP, considering the country's contemporary spaces.

### 3 SÃO TOMÉ'S AND PRÍNCIPE'S PORTUGUESE VARIETIES

In STP, Portuguese is used by approximately 98.4% of the population, a high proportion of speakers in contrast to the national languages, spoken by small groups of users: Santome (36.2%), Lung'le (1.0%), Angolar (6.6%), and Kabuverdianu (7.86%) (INE, 2012). Although the 2012 census does not provide information about the L1 or L2 status of these languages, the percentage distribution provided by the document allows us to infer an imbalance in the coexistence of the official language with other languages of the archipelago.

Portuguese acquisition is a prerequisite for education, employability, and access to different means of communication in STP (ARAUJO; AGOSTINHO, 2010). Portuguese is the main language used in urban centers, which is a result from its transition from L2 to L1 (GONÇALVES, 2010). The irregular linguistic transmission identified in Portuguese dissemination in STP results in structural reanalyses (LUCCHESI; BAXTER, 2009) that, consequently, ascribe identity/linguistic features to the Portuguese varieties that have emerged in the archipelago. It is therefore natural that some traits of the native languages, in the condition of L1, were transported and used in Portuguese L2, which were later passed on to younger generations and acquired as L1.

Portuguese varieties that have recently emerged in the archipelago have also been influenced by regional, cultural, and social issues. Over time, these aspects have changed the Portuguese spoken in STP, modifying it in relation to other similar varieties, that is, shaping it to the linguistics ecology of the islands and to speakers' reality. Authors such as Gonçalves (2010), Bouchard (2017), Gomes (2018), and Santiago and Agostinho (2020), among others, discuss some of the peculiarities of Santomean and Príncipe Portuguese, which may or may not share structures with European Portuguese (EP), the linguistic and target norm of the islands' population.

<sup>2</sup> From a linguistic point of view, Cape Verdeans and Angolares did not distance themselves from their group as strongly as Santomeans and Príncipeans. Both Kabuverdianu and Angolar have greater vitality when compared to Santome and Lung'le, as they are used more often and are still spoken as L1. In Príncipe, Kabuverdianu is more widely spoken than Lung'le, the local language. In Rio Grande community, there are monolingual speakers of Angolar and children whose first contact with Portuguese takes place when they start school.

<sup>3</sup> Nomenclature used by Miguel (2008) for Angolan Portuguese (PA) ("*angolanizou-se*", or "it got Angolonized").

Despite referring to PST and PP as urban varieties of STP, these are not the only ones in the archipelago. Other varieties can be heard, such as Portuguese of Almojarife (PA), spoken in the eastern coastal zone of São Tomé Island; Portuguese of Tongas (PT), used in the community of the Roça Monte Café, the most central region of São Tomé Island (cf. BAXTER, 2018); and even varieties that have not yet been registered, such as Angolar Portuguese (PAG), spoken in the Angolar community and in direct contact with Angolar, and Cape Verdean Portuguese (PCV), in contact with Kabuverdianu and spoken by the descendants of contractors. These varieties are geographically separated and have a different frequency of contact with the multiple Creole languages, being inserted in different linguistic microsystems in STP, which in their turn can culminate in the development of distinct linguistic structures.

PST and PP share similarities with other varieties, such as BP and EP, and have unique characteristics that may bring them closer (or not) to Santome, Lung'le, or other languages spoken in the archipelago — linguistic facts that reinforce their status of effective varieties of Portuguese as L1 (BALDUINO, 2018). The phonetic and phonological arguments that reinforce the idea of consolidating PST and PP as identity Portuguese varieties are discussed in Section 5, following the outline of corpus and methods, which figures in Section 4 below.

#### 4 CORPUS AND METHOD

We used two different methods of data collection in order to analyze some general phonotactic properties of PST and PP. To compile the phonological phenomena in STP varieties, we worked with spontaneous speech data collected *in loco* through 60-minute interviews conducted with subjects of different genders and ages. Subsequently, we collected data through tests that included data elicitation through images and repetition of the vehicle phrase “I speak X softly”, where X was replaced by the target item. The first step was key to identify the phenomena analyzed and to determine possible linguistic variables involved in such processes. The second step was critical to examine and test the preliminary hypotheses.

The recording of interviews and controlled speech data took place in closed environments, such as rooms in local buildings. Prior, we requested subjects to fulfill a form to provide us with personal information that could be useful for linguistic analysis.<sup>4</sup> For analytical purposes, the first 15 minutes of each interview were discarded.

Controlled speech recordings were collected in order to test potential linguistic patterns and domains that could motivate the phenomena (ALBANO, 2017). The recording of controlled sentences allowed us to observe the use of lexical items containing the possibility of sound alteration, as in vocalization and nasalization, within the vehicle phrase “I speak X softly”. This sentence allowed us to isolate and examine the target structures — we controlled adjacent segments, lexical and phrasal prominences, and the occurrence of the phenomenon considering its position within the word.

Tables 1–2 show the evaluated data. In total, we collected data from spontaneous speech of 20 subjects and from controlled speech of 12 subjects. They were distributed according to variety, age, gender, and education. Education level was broken down as follows: Medium (6–9 years), high (10–12 years), and higher education (university studies). As explained, the interviews have greater variability in age and education in order to represent a more heterogeneous speech sample of the examined varieties. However, data obtained through elicitation and repetition of vehicle phrases are more homogeneous, with all informants having a high level of education and being between 18–23 years-old.

---

<sup>4</sup> After the consent form was presented and signed, this form was presented at the beginning of each recording session. Questions include: (i) Full name; (ii) Birth date; (iii) Education level; (iv) Country of birth; (v) Parents' country of birth; (vi) Do you speak other languages?; (vii) Do you understand other languages?; and (viii) Have you ever left São Tomé and Príncipe? For how long?

**Table 1:** Subjects – Spontaneous Speech

PST			PP		
Age	Sex	Education	Age	Sex	Education
18	M	High	20	F	Mid
18	F	High	20	F	High
19	F	High	25	F	Mid
19	M	High	26	F	Mid
20	F	Mid	27	M	High
23	F	Mid	31	M	Mid
24	M	High	35	M	Mid
41	M	High	37	F	High
43	M	High	46	F	High
52	F	Higher	50	F	High

Source: Authors' own work

**Table 2:** Subjects – Controlled Speech

PST			PP		
Age	Sex	Schooling	Age	Sex	Schooling
18	M	High	18	M	High
18	F	High	18	F	High
18	F	High	18	F	High
19	M	High	19	M	High
22	F	Mid	20	F	High
23	M	Mid	20	M	High

Source: Authors' own work

All collected data were analyzed in Praat, so we can acoustically determine the occurrence of the analyzed phenomena. The target lexical item was cut and tagged into a smaller WAV audio file to analyze linguistic criteria such as: Target, trigger(s), and segmental and suprasegmental context(s) that could affect the examined phenomena. The analyzed phenomena are displayed in Section 5.

## 5 PHONETIC AND PHONOLOGICAL ASPECTS OF PST AND PP

This Section discusses phonological phenomena observed in PST and PP. Subsection 5.1 briefly addresses vowel nasalization. Subsection 5.2, in turn, addresses rhotic segments. Subsequently, vocalization is the subject of Subsection 5.3, followed by the observation of segmental deletion processes (Subsection 5.4).

### 5.1 VOWEL NASALIZATION

PST and PP have nasalized vowels due to different processes of regressive vowel nasalization: Tautosyllabic nasalization, triggered by a nasal coda, and heterosyllabic nasalization, triggered by a nasal onset (BALDUINO, 2018; ARAUJO; BALDUINO, 2019).

Tautosyllabic nasality/nasalization is attested in stressed (1.a) and unstressed (1.b) syllables, being optional in word boundaries, as seen in (1.c) and (1.d).

- (1) a. **grande** [ˈgrẽ̃n.di] ~ [ˈgrẽ̃.di] ‘big/high/large’  
 b. **grandeza** [grẽ̃n.ˈde.zɐ] ~ [grẽ̃.ˈde.zɐ] ‘greatness/grandeur/hugeness/magnificence’  
 c. **jovem** [ˈʒɔ.vẽ̃] ~ [ˈʒɔ.vẽ̃j] ~ [ˈʒɔ.vɪ] ‘young’  
 d. **homem** [ˈɔ.mẽ̃] ~ [ˈɔ.mẽ̃j] ~ [ˈɔ.mɪ] ‘man’

Nasalization can also be accompanied by diphthongization ([ẽ̃] > [ẽ̃j]). The optionality of nasalization in (1.c) and (1.d) is accompanied by /N/ coda deletion, which occurs before nasality spreading and blocks the process (ARAUJO; BALDUINO, 2019). Coda deletion at word boundary is recurrent in Portuguese varieties spoken in STP (BALDUINO; VIEIRA; FREITAS, 2020) and is a lexical phenomenon common to all PST segments that are licensed at coda level.

According to Balduino (2018), final nasal coda deletion occurs in unstressed syllables. Such deletion does not happen in structures such as **maçã** \*[ma.ˈsa] ‘apple’, that is, the nasalization process is always observed in final stressed syllables. Therefore, stress and syllable seem to have a key role in determining the implementation of this phenomenon.

Regarding the nasality triggered by /ɲ/, we note its optionality, as indicated in the examples in (2) (BALDUINO, 2020):

- (2) a. **banho** [ˈbẽ̃.ɲɔ] ~ [ˈba.ɲɔ] ‘shower’  
 b. **banheira** [bẽ̃.ˈɲe.rɐ] ~ [ba.ˈɲe.rɐ] ‘bath’

In PST and PP, unlike in BP and EP, /ɲ/ may or may not spread its [nasal] feature to the preceding vowel, either in a stressed syllable (2.a) or in an unstressed syllable (2.b). The absence of nasalization, triggered by the palatal nasal, allows us to question a possible ambisyllabic nature of this segment, since, unlike BP, it does not promote mandatory nasality, one of the arguments used in literature to justify /ɲ/ ambisyllabicity (WETZELS, 1997; 2000; 2007).

Finally, nasality optionality is also observed in heterosyllabic nasality. This phenomenon is only triggered in syllables that carry foot prominence, as indicated in (3) (ARAUJO; BALDUINO, 2019):

- (3) a. **cama** [ˈkẽ̃.mɐ] ~ [ˈka.mɐ] ‘bed’  
 b. **camarada** [ˌkẽ̃.mɐ.ˈra.dɐ] ~ [ˌka.mɐ.ˈra.dɐ] ‘comrade’  
 c. **camada** [ka.ˈma.dɐ] ‘layer’

Data in (3) show that, although heterosyllabic nasality is optional even in stressed syllables (3.a), it never occurs in unstressed syllables that do not also bear foot prominence (3.c). The impossibility of nasalization in unstressed syllables, as well as their optionality in stressed syllables, brings PST and PP closer to Santome and Lung’Ie, since, as it is the case in these languages, the nasal



onset does not necessarily trigger nasality (BANDEIRA, 2017; AGOSTINHO; BALDUINO; ARAUJO, 2020). Therefore, further epistemological studies are necessary to analyze the potential impact of language contact in the development of PST and PP phonological system. We will not address this topic here, but we would like to emphasize the importance of the multilingual environment (and its specificities) surrounding PST and PP for the systematization of STP varieties of Portuguese.

## 5.2 RHOTICS

In PST and PP, the group of rhotics has been object of discussion in studies such as Agostinho (2016), Brandão et al. (2017), Bouchard (2017; 2018), Balduino (2019), Brandão (2018), Agostinho, Soares and Mendes (2020), Agostinho and Mendes (2020), and Vieira and Balduino (2021). In such varieties, rhotics are often — albeit not exclusively — produced as a tepe [r] or as a fricative or vibrant uvular [ʁ] and [ʀ]. Distinctly from BP and EP, rhotics do not only alternate in coda, but also in complex onsets, where they occupy a C2 position, as seen in (4.a). Rhotics also alternate in an intervocalic context, a variation that produces homophone homonyms such as (4.b) and (4.c).

- (4) a. **fraco** [ˈfɾa.ko] ~ [ˈfɾa.kʊ] ~ [ˈfra.kʊ] ‘weak’  
 b. **caro** [ˈka.ʁo] ~ [ˈka.ʀo] ~ [ˈka.rʊ] ‘expensive’  
 c. **carro** [ˈka.ʁo] ~ [ˈka.ʀo] ~ [ˈka.rʊ] ‘car’

According to Bouchard (2017), in PST, age group is a relevant variable to determine the use of strong-R — to the author, a uvular fricative [ʁ]. It is more common among younger generations, more specifically among people under 35 years-old, which is equivalent to approximately 80.8% of the archipelago’s population (WORLDBANK, 2019). Older generations, which make up about 19.2% of the population of STP (WORLDBANK, 2019), do not only prefer the weak-R, understood as [r], but also disfavor the use of [ʁ], considering it an inadequate trait of Santomean variety, as it sets PST and EP apart.

Assuming, then, that the variation between weak-R and strong-R: (i) Is observed in different syllables; (ii) Is present in lexical and post-lexical rules; (iii) Is not determined by any segmental context; (iv) Can be identified in the speech of the same speaker; and (v) Is favored by age factor (cf. BOUCHARD, 2017), admitting /r/ and /ʁ/ as phonemes or just postulating the existence of a neutralized rhotic phoneme /R/ do not explain the use of rhotics in PST and PP. Let’s examine different possibilities for the phonological statute of the rhotic.

At first sight, in view of works such as Bouchard’s (2017), which corroborates younger population’s favorability towards the use of [ʁ] in São Tomé, it is possible to assume the existence of two licensed rhotics (strong-R and weak-R) in the phonological charts/inventories of both varieties. These possible phonemes are not able to drive opposition from minimal pairs nor to characterize a state of neutralization /R/, but they would constitute concurrent forms of an ongoing process of sound change. Therefore, they would alternate in all phonotactic contexts in which the rhotic is predicted, including in the speech of the same speaker. In this case, although there are competing forms, the phonological systems of PST and PP would move towards the existence of only one rhotic.

Agostinho, Soares, and Mendes (2020) strengthen this hypothesis. They address the fusion of quasi-phonemes in PP, that is, segments that display intermediate phonological relations, varying in degree of predictability and contrast (AGOSTINHO; SOARES; MENDES, 2020). According to their experiment, there is evidence to suggest fusion and loss of contrast of rhotics in PP, possibly due to: (i) The quasi-phonemic status of strong-R and weak-R, since, in BP, EP, and possibly in past Portuguese varieties taken to the island, the only contrast established by strong-R [r, h, x, ʀ, ʁ] and weak-R [r] is the intervocalic context; (ii) Their low functional load in the system: As demonstrated in their study, strong-R and weak-R produce few phonemic contrasts compared to other language segments, such as /t, d/, /p, and b/, among others; and, finally, (iii) Linguistic contact with Creole languages, especially Lung’Ie (AGOSTINHO; SOARES; MENDES, 2020).

For these authors, although the phonemic contrast between rhotics in 15<sup>th</sup> century Portuguese is still observed in varieties such as BP and EP, it was lost after the contact with creoles of Portuguese lexical base, such as Lung'le, Santome, Angolar, and Kabuverdianu, languages that do not contrast rhotics. In order to support the relevance of the contact between Portuguese and local languages for the loss of contrast between rhotics, Agostinho, Soares, and Mendes (2020) make a parallel among PP and other varieties of Portuguese in which the weakening of the contrast in rhotics is also reported, as Mozambican Portuguese (PM) and some varieties of Brazilian Portuguese (BP) (cf. AGOSTINHO; SOARES; MENDES, 2020).

According to Brandão and Paula (2018), in PM the weakening of contrast between rhotics can be a result of the variety's contact with several Bantu languages, such as Changana, which presents only an alveolar trill /r/. However, in this study, the authors highlight the fact that, in Mozambique, only 10% of the population speak Portuguese as L1. Thus, further studies are needed to analyze the production of rhotics in PM in light of the contrast between speakers of Portuguese as L1 and L2 (BRANDÃO; PAULA, 2018).

Still according to Agostinho, Soares, and Mendes (2020), the contact of Brazilian varieties with Italian dialects, taken to the country in the 19<sup>th</sup> century, led to a loss of contrast between rhotics. This is reflected in the prevalence of intervocalic *tepe* in detriment of other rhotic variants attested in other Brazilian varieties. As demonstrated by Comiotto and Margotti (2019), the loss of contrast between rhotics in these Brazilian varieties is both a linguistic and cultural mark carried by descendants of Italian immigrants from the Veneto region (COMIOTTO; MARGOTTI, 2019).

A common trait between the Portuguese varieties mentioned by Agostinho, Soares and Mendes (2020) is the long-lasting contact with languages with only one rhotic in their phonological system. The authors conclude that quasi-phonemic contrasts of low functional load/role tend to merge in situations of deep contact with languages with just one rhotic. Thus, their experiment focused on PP is in line with the sociolinguistic analysis undertaken by Bouchard (2017) for PST. Although these studies use different methodological approaches, they point to the same horizon/conclusion: There is a process of ongoing change whose result is the fusion of strong-R and weak-R, and linguistic contact and young speakers' favorability towards the [ʀ] variable are relevant to this process.

Studies such as Bouchard's (2017) and Agostinho, Soares, and Mendes' (2020) demonstrate that the debate around the status of rhotic(s) in the phonological inventory of PST and PP take a different turn than that around the rhotic in BP and in EP, a topic that is subject of controversy. For these varieties, the rhotic status in Portuguese phonological inventory is not consensual — it can be either considered a single phoneme or as two phonemes (CÂMARA JR., 1953; 1970; 1971; MATEUS; D'ANDRADE, 2000; ABAURRE; SÂNDALO, 2003).

When analyzing the synchronic consonant framework of PST and PP, Balduino (unpublished) assumes only one rhotic phoneme, like Agostinho, Soares, and Mendes (2020). However, following the studies of Câmara Jr. (1953) and Abaurre and Sândalo (2003), the author raises the hypothesis of an abstract analysis, in which both forms [r] and [ʀ] could derive from a past form /r/. In this case, the change process is related to the loss of the adjacency of two identical phonemes.

According to Abaurre and Sândalo (2003), in BP as well as in other Iberian languages such as Spanish (HARRIS, 2002), strong-R corresponds to an epiphenomenon of two identical phonemes, that is, it is an effect of the Obligatory Contour Principle (OCP) (McCARTHY, 1986). Therefore, considering OCP as a natural phenomenon, it is very common for languages to avoid identical adjacent segments, or even adjacent segments whose place of articulation is the same. Thus, corroborating the hypothesis of Câmara Jr. (1953), Abaurre and Sândalo (2003) assume that, in Portuguese, the trill could occur intervocalically geminated, while the *tepe* is interpreted as a positional variant that, due to the loss of the continuity feature, is weakened in an intervocalic environment. In other words, when not geminated, the rhotic is produced as a *tepe* between vowels, whereas, when phonologically geminated, it is performed as a strong-R. In the beginning of a word not preceded by a vowel, the rhotic remains a trill, or it is performed as a fricative or glottal trill in most Brazilian varieties (cf. ABAURRE; SÂNDALO, 2003).

On the one hand, considering that the intervocalic context does not determine the production of a strong-R or weak-R in PST and PP, we cannot assume that there is an ongoing positional weakening process in such varieties. Furthermore, /r/ and /ʀ/ alternate in

all stress, syllable, and segment environments. Therefore, for such varieties, if there is no positional weakening that can be structurally justified, it is not possible to assume the existence of phonologically geminated rhotics that are not superficialized in the output. On the other hand, /r/ is the phonological form that enables a more economical derivation of features for [r] and [ʁ], accomplished by, respectively, switching off/delinking the [continuous] feature and by the posteriorization/backness of the vibrant one (cf. ABAURRE; SÂNDALO, 2003). When evaluated together and based on ABAURRE and SÂNDALO's (2003) studies regarding BP, these arguments lead Balduino (unpublished) to suggest that the change process attested by Bouchard (2017) is triggered by the disappearance of the geminated form /rr/ in PST and PP. Due to their synchronic inexistence, the geminated sequence of rhotics does not require the action of the OCP and, therefore, allows the same derivation for /r/, which can be done from strong or weak-R (currently competing forms).

Thus, phonological forms that would supposedly be /karro/ > ['ka.ʁo] (with OCP implementation) and /karo/ > ['ka.ro] (with intervocalic weakening) in BP and EP are underlyingly identical in PST and PP: /karo/ > ['ka.ʁo] ~ ['ka.ro] and /karo/ > ['ka.ro] ~ ['ka.ʁo]. They are also able to derive the other forms of rhotics used in the language, even in the coda. Together with the fact that [r] is favored by older generations and [ʁ] by the younger population (cf. BOUCHARD, 2017), this suggests that both forms — and their diverse variants — are competing forms derived from /r/.

So far, despite having different proposals and theoretical frameworks, revisited studies show that the rhotic is interpreted as a single phoneme in the phonological framework of PST and PP. In fact, we were not able to verify the opposition between [r] and [ʁ] but did verify the alternation [r] ~ [ʁ] in stressed and unstressed contexts and in different syllabic positions where the rhotic is licensed. This is summarized in Table 1. Thus, there are no minimal pairs opposing strong-R and weak-R and, even in a complex onset, the rhotic varies between tepe and uvular segments, which seems to show that there is no context in which one of the rhotic variants is preferred.

**Table 1:** Rhotic distribution in PST and PP: Syllable and stress

Syllable		Stressed		Unstressed
CV	Parede	[pa.'re.di] ~ [pa.'ʁe.di] 'wall'	Caro	['ka.ro] ~ ['ka.ʁo] 'expensive'
CCV	Prato	['pra.tɔ] ~ ['pʁa.tɔ] 'plate'	Profeta	[pro.'fɛ.tɐ] ~ [pʁo.'fɛ.tɐ] 'prophet'
CVC	Parte	['pa.r.ti] ~ ['paʁ.ti] 'part'	Pergunta	[per.'gũ.tɐ] ~ [peʁ.'gũ.tɐ] 'question'
CV	Rato	['ra.tɔ] ~ ['ʁa.tɔ] 'rat'	Ralado	[ra.'la.dɔ] ~ [ʁa.'la.dɔ] 'grated'
VC	Urso	['ur.sɔ] ~ ['uʁ.sɔ] 'bear'	Urbana	[ur.'ba.nɐ] ~ [uʁ.'ba.nɐ] 'urban'
CVC	Arroz	[a.'roz] ~ [a.'ʁoz] 'rice'	Raspado	[raf.'pa.dɔ] ~ [ʁaf.'pa.dɔ] 'grazed'

Source: Author's own work

The same is observed in relation to the segmental context, whose variation [r] ~ [ʁ] was spectrally noted before all seven vowels of the PST and PP vowel systems. Table 2 displays examples of this alternation, considering only CV and CCV syllables where rhotic alternation is not expected in varieties such as PB and PE.

**Table 2:** Rhotic distribution in PST and PP: Vowel and stress

Vowel		Stressed		Unstressed
[i]	Carrinha	[ka.'rĩ.ɐ] ~ [ka.'ʁĩ.ɐ] 'van'	Perigoso	[pe.ri.'go.zɔ] ~ [pe.ʁi.'go.zɔ] 'dangerous'
[e]	Careta	[ka.'re.tɐ] ~ [ka.'ʁe.tɐ] 'grimace'	Reprovar	[re.po.'va] ~ [ʁe.po.'va] 'to fail'

[ɛ]	Carece	[ka.'rɛ.sɪ] ~ [ka.'ʁɛ.sɪ] 'lacks'	Preguiçoso	[prɛ.gi.'so.zʊ] ~ [pʁɛ.gi.'so.zʊ] 'lazy'
[a]	Tirar	[ti.'ra] ~ [ti.'ʁa] 'to remove'	Tira	[ 'ti.rɐ] ~ [ 'ti.ʁa] 'removes'
[ɔ]	Prova	[ 'prɔ.vɐ] ~ [ 'pʁɔ.vɐ] 'test'	Roçado	[rɔ.'sa.dʊ] ~ [ʁɔ.'sa.dʊ] 'mowing'
[o]	Arroz	[a.'rɔʒ] ~ [a.'ʁɔʒ] 'rice'	Provado	[pro.'va.dʊ] ~ [pʁo.'va.dʊ] 'proved'
[u]	Arruma	[a.'ru.mɐ] ~ [a.'ʁu.mɐ] 'tidies up'	Caro	[ 'ka.rʊ] ~ [ 'ka.ʁʊ] 'expensive'

Source: Author's own work

In view of the analyzed data, we corroborate previous studies by Agostinho (2016) and Agostinho, Soares, and Mendes (2020) that indicate the existence of a single phoneme in the consonant inventory of PST and PP. However, it is not our intention to provide a definitive answer to the topic, as it is possible that the rhotic is in process of change (cf. BOUCHARD, 2017) in PST and PP, a fact that is difficult to specify without examining social variables more closely. Furthermore, it must be considered that rhotic alternation does not occur categorically in the speech of older and younger speakers. Although older age groups favor the use of [r] whereas the younger population favors [ʁ] (BOUCHARD, 2017), both forms are identified in the speech of both groups. The discussion around the theme is vast. For now, our intention is to bring the issue to light, indicate the behavior of the data examined, and review the relevant literature. However, it is important that the issue be evaluated in light of new studies.

### 5.3 VOCALIZATION

In this process, a consonant loses its consonantal feature and acquires the characteristics of a vowel for different segmental reasons. In PST and PP, vocalization targets the coda /l/ and the consonant /ɲ/, produced as [w] and [j], as shown in Table 3.

Table 3: Vocalization in PST and in PP

	/l/ vocalization	/r/ ([ʁ] > [w]) vocalization	/ɲ/ vocalization
[i]	fácil [ 'fa.si <u>w</u> ] 'easy'	-----	senhor [sĩ.'jo] 'mister'
[e]	potável [pɔ.'ta.ve <u>w</u> ] 'potable'	ferver [few.'ve] 'to boil'	banho [ 'bɛ̃.jʊ] 'shower'
[ɛ]	mel [ 'mɛw] 'honey'	perto [ 'pɛw.tʊ] 'close'	-----
[a]	salgada [saw.'ga.dɐ] 'salty'	amargo [a.'maw.gʊ] 'bitter'	ganhar [gɛ̃.'ja] 'to win'
[ɔ]	anzol [ɛ̃.'zɔw] 'fishhook'	-----	-----
[o]	bolso [ 'bow.sʊ] 'pocket'	-----	sonho [ 'sõ.jo] 'dream'
[u]	-----	-----	unha [ 'ũ.jɐ] 'nail'

Source: Author's own work

Data in Table 3 show that /l/ can be either realized as [w] in coda, as seen in the first column, and as a velar lateral [ʎ] (BALDUINO; VIEIRA, 2020). According to Balduino and Vieira (2020), vocalization of /l/ occurs before all vowel qualities, except [u] — favorable context for coda deletion — and brings STP varieties closer to Brazilian varieties (QUEDNAU, 1993). However, despite the similarities observed, while the phenomenon is registered as a diatopic process in BP (cf. CÂMARA JR., 1970; MATEUS; D'ANDRADE, 2000), oscillation between [ʎ] ~ [w] occurs recurrently in the speech of the same speaker, as in **fácil** [ 'fa.siw] ~ [ 'fa.siʎ].<sup>5</sup>

Furthermore, [w] is also a rhotic variant, as seen in the second column of Table 3. In the examples **ferver** [few.'ve] 'to boil', **perto** [ 'pɛw.tʊ] 'close' and **amargo** [a.'maw.gʊ] 'bitter', vocalization is fed by rhotic lambdacism in coda, a recurring phenomenon in PST,

<sup>5</sup> All occurrences in Table 3 contained the following variation: [ʎ] ~ [w] for the lateral in coda; [r] ~ [ʁ] ~ [ʎ] ~ [w] for cases in which the rhotic was vocalized and [ɲ] ~ [j] when the palatal nasal was performed as a glide.

but has not been reported in PP so far. So, firstly, there is the realization of rhotic as a velarized lateral, as in **amargar** [a.maʁ.ˈgaɾ] ‘bittering’, feeding the vocalization **amargar** [a.maʋ.ˈgaɾ] ‘bittering’ at a later stage.

Finally, [ɲ] can still be performed as [j], but it then maintains nasalization of the contiguous vowel on the left as shown in the last column of Table 3. Distinctly from BP, [j] is performed as an oral vocalized consonant and is relocated to the onset position while the nasal consonant is eliminated. In these cases, nasalization only remains in the syllabic vowel on the left. Furthermore, albeit less frequent, occurrences such as **ganhar** [ga.ni.ˈaɾ] ‘to win’ were observed in the PST data. Combined with [ɲ]-vocalization, this leads us to question the phonological status of /ɲ/ in such varieties. As in the rhotic case, there is the hypothesis that this is going through a change, because great instability is observed around this segment. Considering, however, that minimal pairs are still observed in both varieties, such as **minha** [ˈmi.ɲɐ] ‘my/mine’ and **mia** [ˈmi.ɐ] ‘meaw’, in which [ɲ] is still produced, we consider /ɲ/ as a phoneme for the purposes of this paper.

In BP, different studies show that the palatal nasal is undergoing variation towards vocalization, or even reduction (ALMEIDA, 2004; PINHEIRO, 2009). In general, these studies indicate both the production of [j], with the maintenance of the nasalization of the preceding vowel, as well as a nasalized anterior vowel: [ĩ]. The segment [ĩ] is also attested in PST and PP, being observed in words whose palatal nasal is preceded by a coronal vowel [i]: **senhor** [si.ˈo] ‘mister’ and **minha** [ˈmi.ɐ̃] ‘my/mine’.

Another common process in STP and PE varieties is the depalatalization of the palatal lateral /ʎ/, performed as [lj] — a coronal lateral followed by a semivowel, as explained in (5). However, we were not able to observe /ʎ/-vocalization into [j], as in **bolha** \*['bo.jɐ] ‘bubble’, in the collected data (BALDUINO, 2020).

- (5) a. **bolha** [ˈbo.ljɐ] ~ [ˈbo.ʎɐ] ‘bubble’  
 b. **milho** [ˈmi.ljɔ] ~ [ˈmi.ʎɔ] ‘corn’

Depalatalization of the lateral palatal is commonly associated with BP (MADUREIRA, 1999; PINHEIRO, 2009). Still, at the beginning of the 20<sup>th</sup> century, Leite de Vasconcelos (1901) already pointed out that the Portuguese spoken in former Portuguese colonies also depalatalized /ʎ/, a fact confirmed by PST and PP data. In EP, [lj] is also verified, constituting a possible variant in the metropolitan region of Lisbon — even though the most recurrent variants are [ʎ] or even the palatalized lateral [lʎ] (QUANDT, 2014).

In short, we noted that PST and PP tend to depalatalize /ʎ/ and /ɲ/, similarly to other Portuguese varieties, while the complete vocalization of /ɲ/ into [j] is also possible. Vocalization also has other targets: The lateral in coda, performed as [w], or even as the rhotic — in case it is lateralized before. In all cases, vocalization and/or depalatalization, configuring weakening processes, alter the syllabic structure of the output, which leads to the following restructuring: CVC \_ CVG (/l/ \_ [w]); C.CV \_ CGV (/ʎ/ \_ [lj]), considering the lateral palatal as ambisyllabic); C.CV \_ GV (/ɲ/ \_ [j] ~ [nj], considering the palatal nasal as ambisyllabic) and C.CV \_ V (/ɲ/ \_ ɹ, deletion of the palatal nasal after nasalization of the previous vowel). In addition to the deletion of /ɲ/, other elisions that cause syllabic restructuring in PST and PP are recurrent. This is the subject of Section 5.4 below.

#### 5.4 SEGMENTAL DELETIONS

Segmental deletions are common in PST and PP, as well as in some varieties of BP and EP. Frequent targets of such phenomenon include: C2 of a cluster (C1C2V), codas, unstressed vowels in pretonic, medial, and final posttonic syllables, as shown in Table 4.

Table 4: Segmental deletions in PST and PP

	Cluster (C <sub>1</sub> C <sub>2</sub> V)	Coda	Vowel
[i]	padre [ˈpa.dɪ] ‘priest’	fácil [ˈfa.sɪ] ‘easy’	escrever [ʃkre.ˈve] ‘to write’

[e]	-----	vermelho [ve.'me.ʎo] 'red'	chávena ['ʃav.nɐ] 'cup'
[ɛ]	-----	certo ['sɛ.tɔ] 'correct'	-----
[a]	plástico [pɐ.'tɔ.ʎi.ko] 'plastic'	vezes ['ve.zi] 'times'	chácara ['ʃa.kɐɾɐ] 'farm'
[ɔ]	petróleo [pe.'tɔ.ʎi] 'petroleum'	farol [fa.'ʒo] 'lighthouse'	abóbora [a.'bɔbrɐ] 'pumpkin'
[o]	problema [po.'be.mɐ] 'problem'	bolso ['bo.sɔ] 'pocket'	-----
[u]	professora [pu.fe.'so.ʁɐ] 'teacher'	pulso ['pu.sɔ] 'wrist'	bicho [bʃi] 'animal'

Source: Author's own work

The deletion processes in Table 4 are just a sample of a phenomenon of wide implementation in PST and PP, in which different targets and triggers are observed and can be explained in different ways (BALDUINO; VIEIRA; FREITAS, 2020; BALDUINO, 2020). In general, according to Gomes (2018), while vowel deletion is conditioned to lexical stress, having pretonic and postonic syllables (final or non-final) as domain, complex onset and coda deletions occur regardless of the target syllable being stressed or unstressed (BALDUINO, 2019, 2020).

According to Gomes (2018), postonic vowel deletion mainly affects mid vowels in PST; however, it is also observed with other vowels: **chácara** ['ʃakɾɐ] ~ ['ʃakarɐ] 'farm'; **abóbora** [a.'bɔbrɐ] ~ [a.'bɔburɐ] ~ [a.'bɔborɐ] 'pumpkin'. According to the author, this process is favored in PST when vowel deletion results in a grammatical cluster, composed, above all, but not exclusively, by an obstruent [p, b, d, t, k, g, f, v] and a sonorant [l, r] (GOMES, 2018), template provided in different studies dedicated to Portuguese. However, an expected syllabic template is not always the result of vowel deletion, since pretonic and postonic high vowel deletions often generate a marked output in Portuguese, which is not expected by syllabification rules. It takes place particularly when the deleted segment is preceded or succeeded by a fricative, as in **escrever** [ʃkre.'ve] 'to write' and **chávena** ['ʃav.nɐ] 'cup', a process also attested in European Portuguese. In both cases, syllabic restructuring promoted by the deletion generates a more complex syllabic template that does not always respect the sonority principle, as in **fraco** ['fɾako] 'weak', in which we have a sequence of fricatives in the cluster.

Paradoxically to the emergence of complex structures, recurrent deletions that target codas (BALDUINO; VIEIRA; FREITAS, 2020) and the second consonant of a complex onset generate CV syllables — a structure considered typologically less marked in natural languages (ZEC, 2007). In this sense, the codas and C2 of a complex onset would coincidentally be targets of processes whose result is precisely the less marked syllable in natural languages.

The results of the deletion phenomenon are different: They can foster syllabic restructuring of a CCV and CVC template into CV, or VC into VG and, at the same time, generate an unexpected and/or more complex template, such as VC in CCCV, as shown in Table 5.

Table 5: Syllable Templates resulting from phonological process in PST and in PP

Word			Original Template	Resulting Template
Escola	[iʃ.'kɔ.lɐ]	~ ['ʃkɔ.lɐ] 'school'	VC	CCV
Estrela	[iʃ.'tre.lɐ]	~ ['ʃtre.lɐ] 'star'	VC	CCCV
Correr	[ko.'re]	~ [ko.'re] 'to run'	CVC	CV

Alface	[aḷ.'fa.si]	~ [aw.'fa.si] 'lettuce'	VC	VG
Prática	['pɾa.ti.kɐ]	~ ['pa.ti.kɐ] 'practice'	CCV	CV

Source: Author's own work

The deletions seem to be triggered by distinct typological issues and with opposite results in the phonological system of the PST and the PP: Consonantal deletion favors a simpler syllabic structure; unstressed vowel deletions emphasize and maximize sound difference between stressed and unstressed syllables (cf. NEVINS, 2012), despite the generation of complex syllables.

## 6 FINAL REMARKS

In this paper, we laid out some historical facts that date back to the first and second periods of Portuguese colonization in STP (15<sup>th</sup>–16<sup>th</sup> and 19<sup>th</sup>–20<sup>th</sup> centuries, respectively). They were responsible for creating a scenario of linguistic contact that, combined with the insular environment of STP, led to the emergence of new languages. It is also in this same historical and social context that Portuguese was chosen as target language of the population in STP, despite the scarcity of language input. We also addressed the historical processes of social interaction in the post-colonial period that fostered Portuguese diffusion and led to the obsolescence of indigenous languages. In this period, we observed that the transmission of Portuguese L1 resulted in the establishment of different varieties of Portuguese spoken in STP, as is the case of PST and PP.

The consolidation of these varieties is reinforced by phonological phenomena inherent to such varieties. Through the presentation of processes such as vowel nasalization, rhotic variation, /l/, /r/ and /ɲ/ vocalization, and different segmental deletions, it was demonstrated that, as in any language, PST and PP have their own structural characteristics which may or may not be shared with different varieties of EP, BP and/or any other Portuguese variety.

In line with some phonetic and phonologic factors of PST and PP, our theoretical discussion encompassed historical and social aspects that reinforce their status as effective varieties of Portuguese as L1. However, more in-depth studies on processes and linguistic aspects of PST and PP are necessary to devise a descriptive grammar of such varieties. In addition to promoting the establishment of a unique STP norm in the archipelago, breaking with pressures still exerted by EP, this would also contribute to the description of yet another variety of Portuguese, revisiting aspects already worked on in other varieties and hence shedding light on old and new issues.

## REFERENCES

ABAURRE, M. B. M.; SANDALO, M. F. S. Os róticos revisitados. In: HORA, D.; COLLISCHON, G. (ed.) *Teoria linguística: Fonologia e outros temas*. Paraíba: Editora Universitária/UFPB, 2003. p. 144-180.

AGOSTINHO, A. L. *Fonologia e Método Pedagógico do lung'le*. 2015. 446f. Tese (Doutorado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2015.

AGOSTINHO, A. L. Róticos em contexto intervocálico no Português da Ilha do Príncipe: fonologia e educação. In: ENCONTRO DA ABECS, 9., Brasília, Universidade de Brasília, 28-30 de novembro, 2016.

AGOSTINHO, A. L.; SOARES, E.; MENDES, M. Merging of quasi-phonemes in contact situations: evidence from rhotics in Principense Portuguese. In: ANNUAL MEETING ON PHONOLOGY: 2020, California, University of California Santa Cruz, 18-20 de setembro de 2020.

AGOSTINHO, A. L.; MENDES, M. A grafia dos róticos intervocálicos no português da Ilha do Príncipe: fusão fonológica e ensino. *Veredas - Revista de Estudos Linguísticos*, v. 24, n. 3, p. 154-176, 2020.

- ALMEIDA, M. M. S. As consoantes do português falado no Vale do Cuiabá. *Signum – Estudos da Linguagem*, Londrina, v.7, p.149-163, 2004.
- ALBANO, E. Fonologia de Laboratório. In: HORA, D.; MATZENAUER, C. L. *Fonologia, fonologias*. São Paulo: Editora Contexto, 2017. p. 169-182.
- ARAUJO, G.; AGOSTINHO, A. L. Padronização das línguas nacionais de São Tomé e Príncipe. *Língua e Instrumentos Linguísticos*, v. 26, p. 49-81, 2010.
- ARAUJO, G.; HAGEMEIJER, T. *Dicionário Santome-Português/Português-Santome*. São Paulo: Hedra, 2013.
- ARAUJO, G.; BALDUINO, A.M. Nasalização Vocálica no Português Urbano de São Tomé e Príncipe. *Diacrítica*, v. 33, n. 2, p. 41-68, 2019.
- ARAUJO, G.; Há uma política linguística para o português em São Tomé e Príncipe?. In: SOUZA, S.; OLMO, F. C. (org.). *Línguas em português – A Lusofonia numa visão Crítica*. Porto: Universidade do Porto Press, 2020. p. 173-197.
- BALDUINO, A.M. *A nasalidade vocálica no português falado em São Tomé e Príncipe*. 2018. 296 f. Dissertação (Mestrado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2018.
- BALDUINO, A.M. Apagamento de /R/ e /S/ em coda no Português Principense. *Papia*, v. 29, n. 1, p. 25-39, 2019. Disponível em: [http://revistas.flch.usp.br/papia/article/view/3366/pdf\\_1](http://revistas.flch.usp.br/papia/article/view/3366/pdf_1). Acesso em: 23 jul. 2020.
- BALDUINO, A.M.; VIEIRA, N.M. Distribuição da lateral /l/ em coda no português santomense. *Estudos Linguísticos*, v. 49, n. 2, p. 594-615, 2020. Disponível em: <https://doi.org/10.21165/el.v49i2.2490>. Acesso em: 20 jul. 2020.
- BALDUINO, A.M.; VIEIRA, N.M; FREITAS, S. A coda no Português Santomense (PST) e Principense (PP): aspectos gerais e processos de apagamento. *Revista da Abralín*, v. 19, n. 1, p. 1-26, 2020. Disponível em: <https://revista.abralin.org/index.php/abralin/article/view/1690>. Acesso em: 15 nov. 2020.
- BALDUINO, A.M. Nasality Triggered by /ɲ/ in two Portuguese of São Tome and Principe. *Diadorim*, v. 22, n.3, 2020, Disponível em: <https://revistas.ufrj.br/index.php/diadorim/article/view/36244/21425>. Acesso em: 15 nov. 2020.
- BALDUINO, A.M. Templates silábicos no Português do Príncipe: processos em coda e ‘prevalência’ de sílabas CV. *Filologia e Linguística Portuguesa*, v. 22, n. 2, p. 247-270, 2020, Disponível em: <https://www.revistas.usp.br/flp/article/view/175052/16950>. Acesso em: 15 nov. 2020.
- BALDUINO, A.M. *Fonologia do português de São Tomé e Príncipe*. 2022. 561 f. Tese (Doutorado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2022..
- BANDEIRA, M. *Reconstrução fonológica e lexical do protocrioulo do Golfo da Guiné*. 2017. 439 f. Tese (Doutorado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2017.
- BAXTER, A. O Português dos Tongas de São Tomé. In: OLIVEIRA, M. D.; ARAUJO, G. (org.). *O Português na África Atlântica*. São Paulo: Humanitas/Fapesp, 2018. p. 297-324.
- BOUCHARD, M. *Linguistic variation and change in the Portuguese of São Tomé*. 2017. 389 f. Dissertation (Doctor of Philosophy) – Department of Linguistics, New York University, New York, 2017.



- BOUCHARD, M. A distinctive use of R as a marker of Santomean identity. *Journal of Belonging, Identity, Language and Diversity*, v. 2, n. 1, p. 6-24, 2018.
- BRANDÃO, S. F. et al. Róticos na variedade urbana do Português de São Tomé. *Papia*, São Paulo, v. 27, n. 2, p. 293-315, 2017.
- BRANDÃO, S. F.; PAULA, A de. Róticos nas variedades santomense e moçambicana do Português. In: BRANDÃO, S. F. (org.). *Duas variedades africanas do Português: variáveis fonético-fonológicas e morfossintáticas*. São Paulo: Blucher, 2018. p. 93-118.
- BRANDÃO, S. F. Apagamento de R em coda externa em duas variedades africanas do português. *Diadorim*, Rio de Janeiro, v. 20, Especial, p. 390-408, 2018.
- CÂMARA JR, J. M. *Problemas de linguística descritiva*. Rio de Janeiro: Editora Vozes, 1971.
- CÂMARA JR, J. M. *Estrutura da Língua Portuguesa*. Rio de Janeiro: Editora Vozes, 1970.
- CÂMARA JR, J. M. *Para o Estudo da Fonêmica Portuguesa*. Rio de Janeiro: Editora Vozes, 1953.
- COMIOTTO, A; MORGOTTI, F. Uso dos róticos do português em contato com os dialetos italianos. *Acta Scientiarum Language and Culture*, v. 41, n. 2, p. 1-9, 2019.
- FERRAZ, L.I. *The creole of São Thome*. Johannesburg: Witwatersrand University Press, 1979.
- GOMES, D. K. Vogais em contexto postônico medial no português de São Tomé. In: BRANDÃO, S. F. (ed.). *Duas variedades africanas do português: variáveis fonético-fonológicas e morfossintáticas*. São Paulo: Blucher, 2018. p. 159-176.
- GONÇALVES, R. *Propriedade de Subcategorização verbal no português de S. Tomé*. 2010. 151f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade de Lisboa, Lisboa, 2010.
- GONÇALVES, R.; HAGEMEIJER, T. O português num contexto multilingue: o caso de São Tomé e Príncipe. *Revista Científica da Universidade Eduardo Mondlane*, v.1, n.1, p. 84-103, 2015.
- HAGEMEIJER, T. 2009. Initial vowel agglutination in the Gulf of Guinea creoles. In: ABOH, E.; SMITH, N. (ed.). *Complex processes in new languages*. Amsterdam/Filadélfia: John Benjamins Publishing Company. 2009. p. 29-50.
- HAGEMEIJER, T. The Gulf of Guinea creoles: genetic and typological relations. *Journal of Pidgin and Creole Languages*, v. 26, n. 1, p. 111-154, 2011.
- HAGEMEIJER, T. From creoles to Portuguese: Language shift in São Tomé and Príncipe. In: LÓPEZ, L. A.; GONÇALVES, P.; AVELAR, J. (Ee.). *The Portuguese language continuum in Africa and Brazil*. Amsterdam/Philadelphia: John Benjamins, 2018. p. 169-184.
- HARRIS, J. Flaps, Trills, and Syllable Structure in Spanish. In: CSIRMAZ, A. et al. (org.). *Phonological Answers, MIT Working Papers in Linguistics*, v. 42, p. 81-108, 2002.
- INE. *Instituto Nacional de Estatística (INE): São Tomé e Príncipe em Números*. São Tomé: INE, 2012. Disponível em: <http://www.ine.st/2012.html>. Acesso em: 20 ago. 2013.
- LEITE DE VASCONCELOS, J. *Esquisse d'une dialectologie portugais e*. Paris: Aillaud & Cie, 1901.

- LUCCHESI, D; BAXTER, A. A transmissão linguística irregular. *In: LUCCHESI, D.; BAXTER, A.; RIBEIRO, I. (ed.). O Português Afro-Brasileiro*. Salvador: EDUFBA, 2009. p. 427-445.
- MADUREIRA, E. Reanálise de alguns aspectos d vocalização da lateral palatal no português. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 8, n. 1. p. 125-145, 1999.
- MATEUS, M. H; D'ANDRADE, E. *The Phonology of Portuguese*. Oxford: Oxford Linguistics, 2000.
- MAURER, P. *Principense Grammar texts, and vocabulary of the afro-portuguese creole of the island of Príncipe, Gulf of Guinea*. London: Battlebridge Publications, 2009.
- MCARTHY, J. OCP effects: Gemination and Antigemination. *Linguistic Inquiry*, v. 17, p. 207-63, 1986.
- MILHEIRO, A. São Tomé e Príncipe e o trabalho do Gabinete de Urbanização Colonial (1944-1974). *In: ACTAS do Colóquio Internacional São Tomé E Príncipe numa perspectiva interdisciplinar, diacrónica e sincrónica*, 2012. p. 87-127.
- MONTEAGUDO, H. A invenção do monolingüismo e da língua nacional. *Gragoatá*, Niterói, n. 32, p. 43-53, 1. sem. 2012.
- NASCIMENTO, A. 2010. *História da Ilha do Príncipe*. Lisboa: Oeiras.
- NASCIMENTO, A. *Atlas da Lusofonia. São Tomé e Príncipe*. Lisboa: Prefácio, 2008.
- NEVINS, A. Vowel Lenition and Fortification in Brazilian Portuguese. *Letras de Hoje*, Porto Alegre, v. 47, n. 3, p. 228-233, 2012.
- NUNES, A. H. Identidade e aprendizagem no crioulo haitiano. *PAPIA*, São Paulo, v. 29, n. 2, p. 81-109, jul./dez. 2019.
- PINHEIRO, N. L. *O processo de variação das palatais lateral e nasal no português de Belo Horizonte*. 2009. 142f. Dissertação (Mestrado em Linguística Teórica e Descritiva) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.
- QUANDT, V. A lateral palatal no Português do Brasil e no Português europeu: um estudo sociolinguístico comparativo. *In: PAULA, A. et al. Uma História de Investigações sobre a Língua Portuguesa: Homenagem a Silvia Brandão*. São Paulo: Blucher, 2014. p.151-166.
- QUEDNAU, L. R. *A lateral pós-vocálica no português gaúcho: análise variacionista e representação não-linear*. 1993. 110f. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1993.
- SANTIAGO, A. M; AGOSTINHO, A. L. Situação linguística do português em São Tomé e Príncipe. *A cor das Letras*, v. 21, n. 1, p. 39-61, 2020.
- SEIBERT, G. Cabo Verde e São Tomé e Príncipe: Ensino Superior e Trajetórias em Portugal. *In: COSTA, A. B. da; FARIA, M. L. de. Formação superior e desenvolvimento: estudantes universitários africanos em Portugal*. Coimbra: Edições Almedina, 2012. p. 209-243.
- SEIBERT, G. Colonialismo em São Tomé e Príncipe: hierarquização, classificação e segregação da vida social. *Anuário Antropológico/2014*, Brasília, v. 40, n. 2, p. 99-120, 2015.
- WETZELS, W. L. The lexical representation of nasality in Brazilian Portuguese. *Probus*, v. 9, p. 1-4, 1997.

WETZELS, W. L. Consoantes palatais como geminadas fonológicas no português brasileiro. *Revista de Estudos da Linguagem*, v. 9, n. 2, p. 5-15, 2000.

WETZELS, W. L. Primary Word Stress in Brazilian Portuguese and the Weight Parameter. *Journal of Portuguese Linguistics*, Ubiquity Press, v.5, n.6, p.9-58, 2007.

VIEIRA, N.M.; BALDUINO, A.M. Apagamento do rótico em coda no Português Santomense (PST): uma análise sociolinguística. *Revista de Estudos da Linguagem*, v. 29, n. 3, 2021.

WORLDBANK. Health Nutrition and Population Statistics. 2019. Retrieved: 20/11/2020. Disponível em: <https://datacatalog.worldbank.org/dataset/s%C3%A3o-tom%C3%A9-and-principe-world-bank-group-country-survey-2019>. Acesso em: 12 ago. 2020.

ZEC, D. The syllable. In: DE LACY, P (ed.). *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press, 2007. p. 161-194.



Received on February 13, 2020. Approved March 22, 2021.