São Tomé and Príncipe's Toponymy

Amanda Macedo Balduino[1], Gabriel Antunes de Araujo[1,2*] e Ana Lívia Agostinho[3]

[1]Universidade de São Paulo, Avenida Prof. Luciano Gualberto, 403, 05508-900, São Paulo, São Paulo, Brasil. [2]Department of Portuguese, University of Macau, University Avenue, FAH, Macau, China. [3]Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil. * Author for correspondence. E-mail: gabrielaraujo@um.edu.mo

ABSTRACT. This study describes the toponymic profile of the Democratic Republic of São Tomé and Príncipe (STP), a former Portuguese colony. Here, we consider (i) its linguistic contact situation, reflected in the toponyms of non-Portuguese etyma (mainly from national languages), and (ii) the structure of the toponymic syntagma in Portuguese. Toponyms are created according to a place's geographic, social, historical, and linguistic reality. Therefore, STP's toponyms are the locus of the Santomean linguistic ecology and memory. Due to its linguistic diversity, the study of Santomean toponymy is transdisciplinary. It contributes to understanding different languages that share São Tomé and Príncipe's ethnicity and linguistic heritage, including multilingual synchronic and/or diachronic conjunctures and the political importance of Portuguese. STP's toponymy expresses a multilingual conjuncture common to the country since the arrival of the Portuguese and (mainly) enslaved Africans from different regions of Africa. Although the Portuguese language and culture have influenced most local toponymy, the Santomean toponymic profile contains endemic linguistic and cultural specificities. Thus, natural phenomena, along with political and historical events, among others, play a role in forming toponyms and characterizing the local toponymy as unique.

Keywords: onomastics; Portuguese; Santome; Lung'ie; African culture.

Toponímia de São Tomé e Príncipe

RESUMO. Este estudo tem como objetivo apresentar uma descrição do perfil toponímico da República Democrática de São Tomé e Príncipe (STP), ex-colônia portuguesa, considerando (i) a sua situação de contato linguístico, refletida em topônimos de étimo não português (principalmente das línguas nacionais) e (ii) a estrutura do sintagma toponímico em português. Os topônimos são criados de acordo com a realidade geográfica, social, histórica e linguística de uma comunidade. Portanto, os topônimos se constituem como o locus da ecologia linguística e da memória. Devido à sua diversidade linguística, o estudo da toponímia santomense é transdisciplinar e contribui para a compreensão das diferentes línguas que partilham a herança linguística e étnica de STP, incluindo conjunturas sincrônicas e diacrônicas multilingues e a importância política atribuída às línguas. A toponímia de STP exprime a conjuntura multilíngue do país desde a chegada dos portugueses e africanos ao arquipélago. Embora a língua e a cultura portuguesa tenham influenciado a maior parte da toponímia local, o perfil toponímico santomense contém especificidades linguísticas e culturais endêmicas. Assim, fenômenos naturais juntamente com eventos políticos e históricos, entre outros, atuam na formação de topônimos e caracterizam a toponímia local como única.

Palavras-chave: onomástica; santome; lung'ie; português; cultura africana.

Introduction

This study presents a pioneering discussion of the official toponymy of the Democratic Republic of São Tomé and Príncipe (STP) by describing the main characteristics of STP's toponyms and toponymic expressions. According to Dick (2002), the toponym involves using certain lexical items or expressions to name the characteristics observed on the site or in the events held there to individualize landforms or traits related to human presence. Toponyms combine different factors to produce a communicative sequence, such as the comprehension of an object named in space, the cognitive perception of its details or constituents, the linguistic representation of the perceived trait, and its denominative manifestation (Dick, 1980). A shared

linguistic element has a toponymic function and thus integrates a motivated process by which it is often possible to relate a name to the geographical area the name designates (Dick, 1998).

Toponyms are related to several physical and anthropocultural aspects of a given linguistic community. Toponyms are linguistic markers of time and place, referring to past eras in which the act of naming took place, and so provide clues about the historical, social, and cultural facts of a community (see Terhi & Jan-Ola, 2007). Thus, toponymic studies also consider extralinguistic resources (Carvalhinhos, 2009). The transdisciplinary nature of toponymy studies can contribute to the understanding of different linguistic strata that structure the vernacular expression of a determined speech community within a circumscribed territory. Therefore, it may reflect synchronous and/or diachronic multilingual conjunctures that characterize the observed area (Carvalhinhos, 1998; Dick, 1992; Nash, 2015; Pedersen, 2017).

The Democratic Republic of São Tomé and Príncipe is a country on the West African coast in the Gulf of Guinea, with a population of approximately 180,000 (INE, 2012)[1]. Initially uninhabited, STP's original colonization favored the emergence of a Creole language (Bandeira, 2017; Bandeira, Araujo, & Finbow, 2021; Ferraz, 1979). Subsequently, isolation, removal of specific groups of speakers from the islands, linguistic contributions of African languages, the creative performance of local speakers, and a constant influx of new parties promoted the speciation of that language. In the 16th century, Santome (ISO 639-3 code: cri) developed in the colonial centers (urban and rural) of São Tomé Island. Angolar (aoa) is the language of the descendants of slaves who fled the mills and formed maroon communities. Other groups of speakers were taken to Príncipe Island and Ano Bom Island, where local conditions contributed to their speciation, giving rise to Lung'Ie (pre) and Fa d'Ambô (fab), respectively (Araujo, 2020b). In addition, Kabuverdianu (kea), the heritage language of the descendants of workers from Cape Verde, who arrived there in the late 19th century and throughout the 20th century, is spoken in the archipelago, mainly in Príncipe Island (Agostinho, 2016). Despite such a multilingual scenario, only Portuguese has enjoyed the status of an official language in STP since Portugal's political emancipation in 1975. According to the National Statistics Institute of São Tomé and Príncipe census, about 98.4% of the population declared themselves as Portuguese speakers (Araujo, 2020a).

Since the arrival of the Portuguese, São Tomé and Príncipe has been a place of ethnic and linguistic co-existence, not always peaceful, between different populations. The local toponymy reflects such diversity in this multilingual scenario. A description of the Santomean toponymic profile can thus contribute to studies on the country's multilingual reality and provide clues about the presence of ethnic groups and their languages. In addition to clarifying the demographic constitution of STP, the study of toponyms can provide information about the social status of the national languages spoken in the country and help raise questions about the linguistic policy.

Despite its linguistic and historical similarities with Portuguese-speaking African countries, we cannot assume that the Santomean toponymic profile is equivalent to a simple reproduction of the toponymic profiles of Portugal or African countries colonized by it. Other linguistic and extralinguistic factors, such as the constitution of ethnic groups and their languages, linguistic contact, and political and historical events, play a role in toponym formation and characterize the local toponymy as unique. Thus, in addition to expected similarities between Santomean and Portuguese toponymies, it is possible that toponyms may be created according to the geographic, social, historical, and linguistic reality of STP. Therefore, STP's toponyms are a privileged locus of the memory and linguistic ecology of São Tomé and Príncipe. Due to the relevance of extralinguistic factors in establishing a site's toponymic profile, this study describes Santomean toponymy in relation to its multilingual reality[2]. The rest of the paper is organized as follows: first, the methodological approach and the corpus are present. Then, an analysis of the Santomean toponymic profile is proposed, followed by the final considerations.

Methodology and corpus

Toponymy research can be (i) qualitative (intensive) or (ii) quantitative (extensive) (Tent, 2015). While the former corresponds to a model centralized mainly on etymology, possible meanings, and the etyma of toponyms, the latter is a broad approach focusing on structural, semantic, and sociohistorical patterns. This study analyzed the official toponymy of São Tomé Island following the second approach. An extensive analysis to provide a general and preliminary description of Santomean toponymy compared with the country's

---

[1] According to the WorldBank (2021), the number of inhabitants was estimated in 215,000.

[2] Consequently, multilingualism and linguistic contact, common realities in STP, are marked by the co-presence of different peoples, which leads to linguistic interaction (INE, 2016).

multilingual scenario was performed. Since this was a pioneering study, the methodological and taxonomic approach adopted was based on the available literature on Brazilian and Portuguese varieties of Portuguese[3].

This methodological approach is supported by the fact that STP is a Portuguese-speaking country like Brazil and Portugal. Therefore, Santomean toponyms are linguistic units connected to the Portuguese world and varieties spoken in other regions[4] (Bonvini, 2008). Furthermore, because STP, Portugal, and Brazil are related by the mother tongue of most of the current population and their shared past, it is possible to hypothesize that toponymic profiles are recurrent and typical to the three countries. That said, the general panorama of Santomean toponymy was established as follows (Carvalhinhos, 2009):

1) a) Collection of toponyms from official information media
   b) Creation of a spreadsheet in which different information was tabulated, such as orthography, phonology, and morphological construction; specific terms; generic terms; and location and geographical circumscription (region, district, neighborhood)
   c) Analysis of the data collected

To collect toponyms, we used data from the most recent census in STP by the National Statistics Institute (INE, 2012). We classified 465 toponyms based on STP's separation into four regions and seven districts. Table 1 shows the number of toponyms per region and district. The corpus collected corresponds to the official geographical classification of the locations, described in the census, including only locations with a relevant demographic concentration. Therefore, the data analyzed do not address all STP toponyms[5].

Table 1. Number of toponyms investigated by district.

São Tomé and Príncipe localities						
Region	North-West		Center-West		South-East		Príncipe
District	Lobata	Lembá	Água-Grande	Mé-Zóchi	Cantagalo	Caué	Príncipe Autonomous Region
Total	69	39	50	163	60	31	47

Source: adapted from INE (2012).

A toponym may consist of two elements, generic and specific (Dick, 1992). The generic element refers to the named geographical entity, while the specific element corresponds to the individualization of a geographic term. Both correspond to a toponymic syntagma and are integrated by agglutination and/or juxtaposition, depending on the structural characteristics of the language (Lima, 2012). Toponyms referring to the Santomean macrotoponymy (district) or related to some elements of its microtoponymy (neighborhoods, gardens,[6] streets) were collected and distributed according to their syntagmatic formation structure.

This study analyzed how STP's toponyms are constituted within a toponymic syntagma. First, the corpus was divided into lexical units in Portuguese and non-Portuguese[7]. This division is essential for a discussion of toponymy as a reflection of the archipelago's complex and multilingual linguistic ecology. It also allows the systematization of data for analyzing the toponymic syntagma and its formation mechanisms. The most recurrent toponymic and generic syntagmata in the STP territory were delineated, observing factors such as pure and toponymized generic items and opacity and transparency (Carvalhinhos, 2009). The two groups were distinguished using dictionaries, grammars, and linguistic works, such as Houaiss and Villar (2001) and (Aulete, 2019) for Portuguese, Araujo and Hagemeijer (2013) for Santome, and (Agostinho, 2016) and Agostinho and Araujo (2021) for Lung'Ie. Finally, a taxonomic grouping was suggested for toponyms based on the taxonomy proposed by Dick (1992).

Toponymy in São Tomé and Príncipe: an overview

This section describes the STP toponymic profile, details the analysis conducted, and demonstrates the results obtained, according to the steps in (2).

---

[3] Thus, this study is related to the *Atlas Toponímico do Brasil/Brazilian Toponymic Atlas* (ATB) (Couto, 2009) and the *Atlas Toponímico de Portugal/Portugal Toponymic Atlas* (ATPor) (Dick, 1980, 1998) projects.

[4] Carvalhinhos (2009) affirms that is essential to know Portugal's toponymy to understand the Brazilian toponymy.

[5] During fieldwork carried out in STP (in 2016 and 2019), speakers sometimes referred to regions using other toponyms, either using lexical units or adding generic items to the main name. In addition, the data from INE present only a few toponyms from Príncipe. However, these issues were not addressed here.

[6] *Roça* is the local term for an agro-industrial rural property.

[7] National Creole languages or another African continent language or family.

2) a) Collection of toponyms per district
    b) Establishment of two distinct toponym language groups
    c) Preliminary examination of toponyms and toponymic syntagmata
    d) Semantic classification of toponyms collected

STP demographic census data lists 465 distinct toponyms, distributed heterogeneously across the country, as shown in Table 2.

Table 2. Percentage distribution of toponyms in relation to the population and territorial area of each district.

District	Toponym	%	Population	Area (km^2)
Água Grande	56	12	69,464	17
Cantagalo	60	12.9	17,161	119
Caué	31	6.7	6031	267
Lembá	39	8.4	14,652	122
Lobata	69	14.8	19,365	229
Mé-Zóchi	163	35.1	44,752	105
Príncipe	47	10.1	7,324	142
Total	465	100	178,739	1001

Source: based on INE (2012) data.

The highest concentration of toponyms is in the Mé-Zóchi district (35.1%), which has an area of 105 km2 and about one-fifth of the population (44,752) (Table 2). Thus, the convergence between space and demography, and not just one of these factors singularly, is a relevant factor for the establishment of São Tomé toponyms, as there is an increase in the need to establish and consequently to name localities, due to the somewhat homogeneous population distribution in each area. We selected a group of toponyms composed of items of non-Portuguese and Portuguese languages was selected by examining each lexical item that individually makes up Santomean toponymy shown in Table 3.

Table 3. Percentage distribution of Santomean toponyms in relation to their etymology.

	Occurrence	%
Non-Portuguese etyma	105	22.6
Portuguese etyma	360	77.4
Total	465	100

Source: elaborated by the authors.

Toponyms of non-Portuguese etyma (22.6%) were those that contained non-Portuguese terms, while toponyms of Portuguese etyma (77.4% of the total) were those that contained items only in Portuguese. Both groups are presented in the following sections, and a semantic classification of the toponyms is analyzed, considering both groups.

Non-portuguese toponyms

The non-Portuguese group was further classified into three subgroups: autochthonous languages, languages of unknown etyma, and mixed-composition autochthonous-Portuguese languages. The toponyms in autochthonous languages correspond to those mentioned in works on STP Creole languages. The toponyms in unknown etyma include items that do not belong to Portuguese and/or are not mentioned in works on autochthonous languages[8]. Finally, hybrid toponyms include compounds from Portuguese and from autochthonous languages. Table 4 specifies these subgroups according to the number of occurrences and their percentage to Santomean toponymy (% general) and concerning the non-Portuguese group (% non-Portuguese).

Table 4. Percentage distribution of São Tomé toponyms of African etyma.

	Occurrences	% Non-Portuguese	% General
Autochthonous languages	50	47.6	10.6
Unknown etyma	35	34.3	7.7
Hybrid	19	18.1	4.1
Total	104	100	22.6

Source: elaborated by the authors.

---

[8] Toponyms: Saton, Quimpo, Tchada, Bandú, Meven Ngay, Bugué, Cassuma, Margão, Quifinda, Quilembá, Rodia, Santy, Zampalma, Zampile, Arribana, Benga, Brigoma, Leça, Predroma, Plancas, Plancas II, Pinquete, Ocacoco, Malanza, Ototo, Maiaua, and Plancas. In this group, the toponyms could come from autochthonous languages or from Portuguese. However, they were not analyzed since their etyma could not be confirmed.

Toponyms in autochthonous languages can be attributed mainly to Santome, although there are toponyms in Lung'Ie in the autonomous region of Príncipe. In (3) and (4), some toponyms formed with lexical items from Santome and Lung'Ie, respectively, are listed[9,10].

3) a) Cabalo Molê   [kaˈbalu moˈle]   lit.[11] 'dead horse'
    b) Ubua Coconja   [ˈubwa koˈkõdʒa]   lit. 'coconut tree fence'
    c) Libadoque   [ˈliba dɔˈke]   lit. 'top of the slope'

4) a) *Ximalô*   [ʃĩmaˈlo]   lit. 'top'
    b) *Lentapia*   [lẽˈta ˈpja]   lit. 'to go in and look'

In (3), the items of the toponymic syntagma were recovered from the *Santome-Portuguese/Portuguese-Santome Dictionary* (Araujo & Hagemeijer, 2013). Whereas in (4), the source of the toponym classification is a glossary of Lung'Ie (Agostinho & Araujo, 2021). In (3b), for example, while *pia* [ˈpja] means 'to look' in Lung'Ie and Santome, *lenta* [lẽˈta] means 'to go in; entrance' (Agostinho & Araujo, 2021, p. 78) in Lung'Ie and Santome (Araujo & Hagemeijer, 2013, s. v.). Considering that Lung'Ie is spoken in Príncipe Island, a region where the toponym *Lentapia* is used, we can argue that *Lentapia* is a toponym constructed from lexical items of Lung'Ie.

Toponyms can be formed by a generic item plus a specific word characterizing a toponymic syntagma or by reduplicated lexical items in the source language, regarding the structure of toponyms in autochthonous languages and according to toponymized common names. In general, toponyms composed of only one common noun are equivalent to a toponymized generic item (5a), an abstract noun (5b), or, more frequently, a reference to the local flora (5d–f).

5) a) *Poto*   [ˈpɔtɔ]   lit. 'door'
    b) *Blublu*   [bluˈblu]   lit. 'hurry'
    c) *Vua Sata*   [ˈvwa saˈta]   lit. 'shortcut'
    d) *Mucambu*   [muˈkɛ̆bu]   '*Vernonia amygdalina*'[12]
    e) *Macamblalá*   [makɛ̆blaˈla]   '*Crasterispermum montanum*'
    f) *Coconja cuto* [koˈkõdʒa ˈkutu]   lit. 'short coconut tree'

In (5a), *poto* lit. 'door' is considered a toponymized generic item because *poto* is regularly found in toponymic syntagmata, as in *Poto Correia* and *Poto Mandioco*. A geographic element can be constituted by elements of nature (Araujo & Hagemeijer, 2013), but it can also come from an artificial geographic landmark. The toponym *Vua Sata* lit. 'shortcut' (a nominal compound in Santome) corroborates this statement. This name refers to a path that shortens the distance between two points, a concept established from two geographical landmarks, the place of departure and the place of arrival, both anthropological concepts.

In (6), some toponymic syntagmata formed by a generic item plus a toponym are presented.

6) a) *Mato Andim*   [ˈmatuɛ̆ˈdʒĩ]   lit. 'andim field'
    b) *Mato Kilemba* [ˈmatu kiˈlẽba]   lit. 'kilemba field'
    c) *Ubua Coconja* [ˈubwa koˈkõdʒa]   lit. 'coconut fence'
    d) *Ubua Budo*   [ˈubwa ˈbudu]   lit. 'stone fence'
    e) *Uba*[13] *Caju*   [ˈuba kaˈʒu]   lit. 'cashew fence'

In (6a,b), *mato* means 'field' in Santome (Araujo & Hagemeijer, 2013, s. v.), a generic item found in toponymic syntagmata on São Tomé Island but not on Príncipe Island, whose form in Lung'Ie is *umatu* (Agostinho & Araujo, 2021). These examples show a nominal construction of possession common to Santome and Lung'Ie. The possession relationship between two names is obtained by concatenating the postponed possessor item to the possessed item. In (6c,e), conversely, the generic term is *ubua/uba* in São Tomé Island and *uba* in Príncipe Island, both meaning 'fence.' As in the case of *poto*, it refers to an element with a cultural nature to refer in a generic way to a geographical area. In this sense, it is possible to construct toponymic

---

[9] The spelling of toponyms was kept as it appears in the data from the 2011 Census (INE, 2012), which does not use the official alphabet *Alfabeto Unificado para as Línguas Nativas de São Tomé e Príncipe* (ALUSTP) for the native languages. See Araujo and Agostinho (2010) for details. In the census report, Portuguese-inspired spelling was used for terms from autochthonous languages.

[10] The phonetic transcription used in this paper is from local Portuguese varieties, since it is the language of 98% of the population, unless noted otherwise.

[11] Because it is a toponymic item used locally as such, the toponym is preceded by the word *literally* (lit.), whenever necessary, to show the relationship between Portuguese and other languages.

[12] The scientific names of fauna and flora are written in italics following the two-name nomenclature standard. See Araujo (2021) for a fauna and flora lexica analysis.

[13] In Lung'Ie: [ˈuɡ͡ba] (Agostinho, Araujo, & Freitas, 2012).

syntagmata in autochthonous languages. The generic item must be a common noun that describes a toponym characteristic features (cf. Dick, 1992).

Reduplication is also adopted as a recurring process of creating Santomean toponyms in autochthonous languages (7a) and Portuguese (7b–d). In the latter, lexical items come from Portuguese, noun reduplication is a grammatical feature of autochthonous languages, while verb reduplication is found in Portuguese (Araujo, 2002).

7)   a)   *Budo-budo*   lit. 'stones'   whereas *budo* [ˈbudu] 'stone'
      b)   *Pema-pema*   lit. 'palm trees'   whereas *pema* [ˈpɛma] 'palm tree'
      c)   *Lama-lama*   lit. 'muddy place'   whereas *lama* [ˈlama] 'mud'
      d)   *Manga-manga*   lit. 'mangoes'   whereas *manga* [ˈmɐ̃ga] 'a mango'

The data in (7) were considered only as toponyms. Although they may be reported in the literature as non-reduplicated forms, all consulted sources did not mention them as such. However, since words like *budo, pema*, *lama*, and *manga* present an independent lexical content, reduplication is a productive process in Santomean. Moreover, in all cases in which the toponym is formed by reduplication, the reduplicated lexical item corresponds to an element of nature. Hence, we hypothesized that a toponym formed by reduplication might indicate a local abundance of the referred natural element. This presupposition must be investigated by analyzing morphological processes in autochthonous languages. In (8), each item of mixed composition contains a non-Portuguese lexical element in the toponymic syntagma.

8)   a)   *Favorita Luchan*   [favoˈɾita luˈʃɐ̃]   lit. 'favorite village'
      b)   *Vaji Grande*   [ˈvaʒi ˈgrɐ̃di]   lit. 'big grotto'/'big field limit'
      c)   *Uba*[14] *Flor*   [ˈuba ˈfloɾ]   lit. 'flower's fence'
      d)   *Obo Coelho*   [oˈbo kuˈeʎu]   lit. 'rabbit forest'
      e)   *Água Izé*   [ˈagwa iˈzɛ]   lit. 'shrimp waters'/'shrimp river'
      f)   *Água Bobô*   [ˈagwa boˈbo]   lit. 'clear waters'/'clear river'
      g)   *Praia Io Grande*   [ˈpɾaja ˈjo ˈgrɐ̃di]   lit. 'beach of Grande River'

The example (8a) has an ambiguous interpretation, as the common noun *luchan* lit. 'village, non-urbanized area' can be a generic item, featuring the toponym *favorite*, 'favorite,' or toponymized together with *favorita*, demarcating a toponym in its entirety. This issue is beyond the scope of this study and will not be addressed. However, the hybrid formation of *Favorita Luchan*, which involves a common name in Santome, *luchan*, and a name from Portuguese, *favorita*, is emphasized.

The examples in (8b–d) indicate a generic element of an autochthonous language (*vaji* lit. grotto/field limit,' *ubua/uba* 'fence,' and *obo* 'forest'), together with a Portuguese lexical item (*grande*, *flor*, and *coelho* for 'huge,' 'flower,' and 'rabbit,' respectively). In (8e,f), we observe the generic *água* 'water,' a productive item in the formation of Santomean toponyms, accompanied by *izé* lit. 'shrimp' and *bobô* lit. 'clear,' terms from Santome. Finally, in (8g), *io* lit. 'river' in Santome is the only autochthonous term in the toponymic syntagma. In this case, *praia* may be the generic element, whereas the compound *io grande* corresponds to the toponym's specific content.

The generic function of the examples in (8) can only be recovered if the speaker is bilingual and knows the lexical items. Additionally, one can access that information through schooling or from bilingual speakers who explain, in some way, the meanings of those terms. Therefore, a lack of knowledge of Santome and its linguistic obsolescence may contribute to its semantic emptying. It also makes lexical items not transparent to monolingual Portuguese speakers who may consider *Favorita Luchan*, *Uba Flor*, *Obo Coelho*, *Ubua Budo*, and *Ubua Conconja*, among others, as full nouns. Consequently, these nouns may be subject to structural and semantic decoding or re-analysis even though it may be possible to recover hybrid toponymic syntagmata using linguistic material or traditional knowledge.

In *Praia Io Grande* (8g), the generic *praia* is added despite the possibility of *io* lit. 'river' also being equivalent to a generic geographical element. This addition may be because *io* is no longer clear to monolingual speakers, who interpret it as an integral part of the toponym. Otherwise, native items can be interpreted as lexemes of Portuguese, leading to sui generis interpretations of the toponymic syntagma, as in *mato* (4a,b). Speakers can decode it as a Portuguese word, although in Santome, it means 'field,' a geographic element distinct from the Portuguese form *mato* lit. 'tall grass.'

Finally, the unknown etyma subgroup includes toponyms such as *Ocaoco*, *Bugué*, and *Cassuma*, which may correspond to items from autochthonous languages that are not dictionarized. It may even have an etymology

---

[14] Variation of *ubua*, from Santome.

of some other continental African language that may have circulated in the archipelago during the colonial period. Table 5 shows the distribution of toponyms of unknown etyma by district.

Table 5. Percentage distribution of toponyms of unknown etyma in relation to districts.

District	Toponym	%
Água Grande	12	24
Cantagalo	11	18.3
Caué	5	16.1
Lembá	8	20.5
Lobata	16	23.2
Mé-Zóchi	45	28.2
Príncipe	7	14.9
Total	104	100

Source: elaborated by the authors.

The highest concentration of toponyms of unknown etyma is in the Mé-Zóchi district (28.2%), a region that also has the highest number of toponyms in São Tomé and Príncipe. Lobata is the second district with the most non-Portuguese toponyms (23.2%), followed by Lembá (20.5%) and Cantagalo (18.3%).

Toponyms in portuguese

Analysis of Portuguese toponymy shows that toponyms can consist of simple names (9a,b), compound names (9c,d), and commemorative dates (9g). It can be formed by a toponymic syntagma comprising a generic and a specific element (9e,f).

9)   a)   *Maracujá*   lit. 'passion fruit'
       b)   *Gratidão*   lit. 'gratitude'
       c)   *Diogo Vaz*   'Diogo Vaz,' a historic proper name
       d)   *Bela Vista*   lit. 'nice view'
       e)   *Água Crioula*   lit. 'Creole water'
       f)   *Quinta das Flores*   lit. 'beach of Grande River'
       g)   *3 de Fevereiro*   lit. 'February 3'[15]

In this study, 38 different generic elements were collected. Table 6 shows the elements according to the number of occurrences recorded in the STP census. There were 162 occurrences of 34 generic elements recorded as Santomean toponymic syntagmata in the 2016 census. Table 6 shows only official data from areas where a population contingent could be found, so there are more occurrences of generic elements in the Santomean territory than the forms specified in Table 6, which will not be addressed here. The toponym *Roça*, for example, is frequently used by Santomeans to designate the old agro-industrial units of the colonial era (1470–1975). It is used within a toponymic syntagma in front of a specific element, as in (10).

10)   a)   *Roça Nunno Oliveira*   lit. 'Nunno Oliveira's farm'
       b)   *Roça Diogo Vaz*   lit. 'Diogo Vaz' farm'

Table 6. Occurrence of generic elements of Santomean toponymy in Portuguese.

Generic	Occur.	Generic	Occur.	Generic	Occur.	Generic	Occur.
Água	29	*Centro*	1	*Quintal*	5	*Pau*	2
Atrás	4	*Ilhéu*	1	*Recta*	1	*Vale*	4
Alto	1	*Estrada*	1	*Rio*	3	*Vila*	20
Angra	2	*Monte*	8	*Ribeira*	2	*Vista*	1
Aldeamento	1	*Monta*	3	*Rua*	9	*Zona*	2
Bosque	1	*Parque*	1	*Riba*	2		
Bairro	2	*Praia*	24	*Roça*	5		
Caminho	5	*Ponta*	7	*Riboque*	6		
Cidade	2	*Ponte*	1	*Rocinha*	1		
Colónia	2	*Porto*	2	*Terra*	1		Total
Campo	1	*Quinta*	6	*Terreiro*	1		162

Source: elaborated by the authors.

---

[15] Reference to the *Batepá Massacre*, which occurred on February 3, 1953, in the capital, when hundreds of natives were killed by the Portuguese colonial administration and landowners (Rodrigues, 2019).

Among the toponymic and toponymic syntagmata analyzed, the establishment of a generic toponymized postponed to the specific term was identified, a construction used as a strategy for disambiguating toponyms, as shown in (11).

11)	a)	*Praia Conchas Roça*	/	Praia Conchas
	b)	*Sundy Praia*	/	*Sundy Praia*
	c)	*Ribeira Palma Praia*	/	*Ribeira Palma*
	d)	*Colónia Açorena Praia*	/	*Colónia Açorena*
	e)	*Cidade Santana Alta*	/	*Cidade Santana Centro*

In (11a–d), the generic elements praia 'beach,' ribeira 'stream,' colónia 'colony,' cidade 'town,' alta 'highlands,' centro 'town center,' and roça 'farm,' when postponed to a specific term, qualify the previous term and constitute a toponymic unit, as in (11e). Therefore, they divest themselves of the generic character, which can be performed optionally by another generic element. The disambiguation of toponyms can also occur by adjectivation of the toponym (12a, b) or even by its numbering (12c,d).

12)	a)	*Guegue*	/	*Guegue Velho*
	b)	*Hospital*	/	*Hospital Velho*
	c)	*Plancas I*	/	*Plancas II*
	d)	*Bairro 1*	/	*Bairro 2*

A second element in the compound suggests a double interpretation for the toponyms in (12). From one side, it can be a stable compound, consisting of two lexical items, such as in *Praia Conchas* (11a) and *Ribeira Palma* (11c). On the other hand, it can also correspond to an item with an added generic part to undo any ambiguity with other toponyms. Nevertheless, these generic parts qualify a specific element, creating a new toponym. Generic toponymization also occurs from isolated elements (13). As discussed in this section, *ilhéu* 'islet' can occur with some toponyms forming a generic syntagma. However, in (13a), *ilhéu* is a toponym and names a neighborhood on its own. *Mesquita*, *Canavial* and *Cachoeira* exemplify cases in which a generic element is adopted as a toponym too.

13)	a)	*Mesquita*	lit. 'mosque'
	b)	*Ilhéu*	lit. 'islet'
	c)	*Canavial*	lit. 'sugarcane plantation'
	d)	*Aeroporto*	lit. 'airport'
	e)	*Cachoeira*	lit. 'waterfall'

Some words of African etyma are described as Brazilianisms or Angolanisms in some Portuguese dictionaries (Table 7). For example, words such as Gonga, Binda, and Muçambo, which happen to be toponyms in São Tomé and Príncipe, are listed by Carvalhinhos (2009) and Aulete (2019) as lexical items in Brazil and Angola. Therefore, it would be necessary to analyze whether these terms were incorporated into Portuguese spoken from African languages or via the Portuguese language of the colonizers in STP. At the same time, it is not clear when they started to be used as toponyms in STP.

Table 7. Toponyms with items defined as Brazilianisms or Angolanisms.

Toponym	Definition by (Carvalhinhos, 2009) and Aulete (2019)
Gonga	1. s. m. || (Angola) a Congo basket made of Mateba straw. Cp. *Gongá* (Bras.). 2. s. f. || (Angola) a certain musical instrument. 3. s. f. || (Bras., Northwest) (pop.) old clothes (usually men's).
Binda	1. n. f. || || African name for pumpkin-sheep. Dried fruits are used to transport liquids, for marimbas, etc. || The bowl made of Binda.
Muçambo	1. n. m. || (Angola) metallic ornament that the indigenous people of Lunda use to tighten their braids.

Source: elaborated by the authors.

Table 8 shows the distribution of toponyms in Portuguese by district.

Portuguese toponyms are more common than non-Portuguese items in all districts. However, the highest concentration of Portuguese etyma toponyms is in the Caué district (83.9%), in the countryside, where Creole languages are more frequently used than in urban areas (83.9%), in the countryside, where Creole languages

are more frequently used than in the more urbanized districts (Bouchard, 2017). This dominance reinforces the presence of Portuguese in different parts of STP, not just in the capital or in urban areas, and shows its nationwide spread (Araujo, 2020b).

Table 8. Percentage distribution of toponyms in Portuguese in relation to districts.

District	Toponym	%
Água Grande	34	68
Cantagalo	48	80
Caué	26	83.9
Lembá	29	74.4
Lobata	52	75.4
Mé-Zóchi	110	67.5
Príncipe	39	82.9
Total	360	100

Source: elaborated by the authors.

Semantic classification of Santomean toponyms: a first approach

In semantic terms, the São Tomean toponyms investigated here can be classified according to their physical or anthropocultural nature. Based on the taxonomic model of Dick (1992), the number of occurrences of toponyms is classified by their physical nature (astrotoponyms, cardinotoponyms, chromotoponyms, dimensiotoponyms, phytotoponyms, hydrotoponyms, lithotoponyms, morphotoponyms, and zootoponyms), as shown in Table 9.

Table 9. Occurrence of physical toponyms in Santomean toponymy.

Group	Occurrence	%	Example
Astrotoponyms			
Stars and celestial bodies	4	2.9	*Cruzeiro*
Cardinotoponyms			
Geographical positions	4	2.9	*Atrás do Cemitério*
Chromotoponyms			
Chromatic scale	3	2.2	*Água Bobô*
Dimensiotoponyms			
Dimensional features	9	6.6	*Praia Io Grande; Portinho*
Phytotoponyms			
Flora names	63	46.3	*Binda; Java; Micondó*
Hydrotoponyms			
Hydrographic features	7	5.2	*Cachoeira; Cova Água*
Lithotoponyms			
Minerals and soil composition	12	8.8	*Budo-budo*
Morphotoponyms			
Geometric shape; composition	13	9.6	*Monte Forte; Praia seca*
Zootoponyms			
Fauna names	21	15.4	*Água Izé*
Subtotal	136	100	
Total	465	29.2	

Source: elaborated by the authors.

About 29% of Santomean toponyms can be semantically related to aspects of a physical nature, either in Portuguese or an autochthonous language. In this group, phytotoponyms (46%), zootoponyms (15%), morphotoponyms (10%), and lithotoponyms (9%) have a higher frequency of occurrence, while astrotoponyms (2%) and chromotoponyms (3%) are less frequent. About 29% of Santomean toponyms can be semantically related to aspects of a physical nature, either in Portuguese or an autochthonous language. In this group, phytotoponyms (46%), zootoponyms (15%), morphotoponyms (10%), and lithotoponyms (9%) have a higher frequency of occurrence, while astrotoponyms (2%) and chromotoponyms (3%) are less frequent. Anthropocultural toponyms make up approximately 58% of the corpus, and they were grouped into 14 different taxonomies here (Table 9). Tables 9 and 10 show Santomean toponymy's semantic aspects, indicating a possible motivating essence. Each group must be individually analyzed to verify whether the taxonomy used is sufficient to outline the toponymic profile of STP or not. In cases of taxonomic ambivalence, one nomenclature may be chosen over another.

Table 10. Occurrence of anthropocultural toponyms in Santomean toponymy.

Group	Occurrence	%	Example
Animotoponyms			
Psychic and spiritual life	53	19.70	*Boa Esperança*
Anthrotoponyms			
Proper names	69	25.7	*Diogo Vaz*
Axiotoponyms			
Titles	5	1.9	*Abade*
Chorotoponyms			
Places	8	2.9	*Bombaim*
Chronotoponyms			
Chronological indication	9	3.4	*Hospital Velho*
Ergotoponyms			
Hydrographic accidents	9	3.4	*Batelo*
Ethnotoponyms			
Ethnic elements	1	0.4	*Água Criola*
Dirrematotoponyms			
Phrase and linguistic statement	1	0.4	*Voz d'América*
Hierotoponyms			
Sacred; religiosity	70	26	*São Miguel*
Historiotoponyms			
Events	2	0.7	*3 de Fevereiro*
Hodotoponyms			
Means of communication	1	0.4	*Oque Ponte*
Poliotoponyms	7		
Items such as village and city		2.6	*Aldeia; Centro*
Sociotoponyms			
Professional activities, workplaces; meeting places for members of society	32	11.9	*Água Doutor; Hospital*
Somatotoponyms			
Metaphor for parts of the human or animal body	2	0.7	*Cabeça Cal*
Subtotal	269	100	
Total	465	57.8	

Source: elaborated by the authors.

Although Dick (1992) model allows a safe semantic interpretation of toponyms, it addresses the Brazilian reality and is subject to double interpretation, depending on its structural composition. Thus, toponyms such as *Nova Estrela* have chronological (nova 'new') and *celestial* (estrela 'star') characteristics, allowing a double reading. Such cases need to be re-examined in conjunction with extralinguistic information, such as historical facts and areal morphology described by the toponym.

Final remarks

Santomean toponymy reflects the country's linguistic history and societies. It comprises toponyms of Portuguese, autochthonous, and continental African language etyma. Sometimes toponyms are hybrid, formed by elements of two or more languages. STP toponymy allows for observation of the country's multilingual reality while revealing the overlap of toponyms in Portuguese and the region's autochthonous languages, thus reflecting practices of naming followed across the world and of power relations (Pedersen, 2017; Terhi & Jan-Ola, 2007). Therefore, STP toponymy expresses the multilingual conjuncture that has accompanied the country since the arrival of the Portuguese and Africans from different regions of Africa.

Among the toponyms described, 77.4% are of Portuguese etyma, while 22.6% are composed entirely or partially of words of non-Portuguese etyma. Data from the latter group were systematized into two subgroups: fully autonomous toponyms (18.2%), such as *Ubua Coconja* [ˈubwa koˈkõdʒa] lit. 'around coconut trees,' and mixed toponyms (4.1%), such as *Favorita Luchan* [favoˈɾita luˈʃɐ̃] lit. 'favorite village.' Although the Portuguese language has heavily influenced local toponymy, the Santomean toponymic profile is not equivalent to a mere transplanted Portuguese toponymic profile. On the contrary, it contains endemic linguistic and cultural specificities. Thus, natural phenomena and local political and historical events, among others, play a role in toponym formation and characterize the local toponymy as unique.

Toponym lexica are created according to the geographic, social, historical, and linguistic reality of a community. Therefore, STP's toponyms are a privileged locus of the linguistic ecology and memory of São Tomé and Príncipe. Hence, toponyms constitute an additional element that helps to understand Santomean

culture. Therefore, the study of Santomean toponymy contributes to the comprehension of different strata that structure the vernacular expression of STP. It also reflects synchronic and diachronic conjunctures of the islands and aspects related to their occupation. Additional studies on the archipelago's toponyms could clarify linguistic and historical issues offering arguments to support the flow of slaves or populations in each specific region. It may also reveal a picture of the languages spoken by different ethnic groups and their inter-relationships.

Acknowledgments

References

Agostinho, A. L. (2016). *Fonologia do Lung'Ie*. Muenchen: LINCOM GmbH.

Agostinho, A. L., & Araujo, G. A. d. (2021). *Lung'Ie, lunge no*. São Paulo, SP: FFLCH/Humanitas.

Agostinho, A. L., Araujo, G. A. d., & Freitas, S. (2012). Resolução de hiato externo em principense. *Papia, 22*, 295-305. Retrieved from http://www.revistas.fflch.usp.br/papia/article/view/1672/1483

Araujo, G. A. d. (2002). Truncamento e reduplicação no português brasileiro. *Revista de estudos linguísticos, 10*, 61-90. DOI: http://dx.doi.org/10.17851/2237-2083.10.1.61-90

Araujo, G. A. d. (2020a). Há uma política linguística para o português em São Tomé e Príncipe? In S. Souza & F. C. Olmo (Eds.), *Línguas em português. A Lusofonia numa visão Crítica* (Vol. 1, p. 173-197). Porto, PT: University of Porto Press.

Araujo, G. A. d. (2020b). Portuguese language expansion in São Tomé and Príncipe: an overview. *Diadorim, 22*, 57-78. DOI: https://doi.org/10.35520/diadorim.2020.v22n1a32012

Araujo, G. A. d. (2021). Expansão do léxico português em São Tomé e Príncipe : fauna e flora. *Études romanes de Brno, 42*(1), 27-55. doi:10.5817/ERB2021-1-3

Araujo, G. A. d., & Agostinho, A. L. (2010). Padronização das línguas nacionais de São Tomé e Príncipe. *Língua e Instrumentos Linguísticos, 26*, 49-81. Retrieved from http://www.revistalinguas.com/edicao26/artigo_3.pdf

Araujo, G. A. d., & Hagemeijer, T. (2013). *Dicionário livre santome/português (Livlu-nglandji santome/putugêji)*. São Paulo, SP: Hedra.

Aulete, C. (2019). *Dicionário Aulete da Língua Portuguesa*. Retrieved from http://www.aulete.com.br/

Bandeira, M. (2017). *Reconstrução fonológica e lexical do protocrioulo do Golfo da Guiné*. München, DE: Lincom.

Bandeira, M., Araujo, G. A. d., & Finbow, T. (2021). The gulf of guinea proto-creole and its daughter languages: from liquid consonants to complex onsets and vowel lengthening. *Journal of Language Contact, 14*(3).

Bonvini, E. (2008). Línguas africanas e português falado no Brasil. In J. L. Fiorin & M. Petter (Eds.), *África no Brasil: a formação da língua portuguesa*. São Paulo, SP: Contexto.

Bouchard, M.-E. (2017). *Linguistic variation and change in the Portuguese of São Tomé.* (Doctor of Philosophy). New York University, New York.

Carvalhinhos, P. (1998). *A toponímia portuguesa: um recorte linguístico do Douro ao Tejo.* (Tese de Doutorado). Universidade de São Paulo, São Paulo.

Carvalhinhos, P. (2009). Projeto Atlas Toponímico de Portugal: Informes iniciais, critérios, recorte: projeto variantes lexicais na toponímia Portuguesa. A questão do genérico. In *Anais do SILEL, 1*, Uberlândia, MG: Edufu.

Couto, H. H. (2009). *Linguística, ecologia e ecolinguística*. São Paulo, SP: Contexto.

Dick, M. V. d. A. (1980). *A motivação toponímica. Princípios teóricos e modelos taxionômicos.* (Tese de Doutorado em Filologia e Língua Portuguesa). Universidade de São Paulo, São Paulo.

Dick, M. V. d. A. (1992). *Toponímia e antroponímia no Brasil* (Coletânea de Estudos). São Paulo, SP: Universidade de São Paulo.

Dick, M. V. d. A. (1998). Atlas Toponímico: um estudo dialetológico. *Revista Philologus, 10*, 61–69.

Dick, M. V. d. A. (2002). Aspectos de etnolingüística– a toponímia carioca e paulistana – contrastes e confrontos. *Revista USP*, 180-191.

Ferraz, L. I. (1979). *The Creole of São Tomé*. Johannesburg: Witwatersrand University Press.

Houaiss, A., & Villar, M. d. S. (2001). Rio de Janeiro, RJ: Objetiva.

Instituto Nacional de Estatística [INE]. (2012). *São Tomé e Príncipe em Números*.

Instituto Nacional de Estatística [INE]. (2016). *Instituto Nacional de Estatística de STP: Dados demograficos 1970 à 2016*. Retrieved from http://shorturl.at/EGPXZ

Lima, A. (2012). *De Bracara Augusta a Braga: análise toponímica de um concelho português.* (Dissertação de Mestrado em Filologia e Língua Portuguesa). Universidade de São Paulo, São Paulo.

Nash, J. (2015). Is Toponomy necessary? *Studies in languages, 39*(1), 228-239. DOI: https://doi.org/10.1075/sl.39.1.08nas

Pedersen, A. (2017). The transmission of toponyms in language shift societies. In A. Terhi & O. Jan-Ohala (Eds.), *Socio-onomastics the pragmatics of names* (p. 22-44). Amsterdam, NL: John Benjamins.

Seibert, G. (2015). Colonialismo em São Tomé e Príncipe: hierarquização, classificação e segregação da vida social. *Anuário Antropológico, 40*, 99-120.

Terhi, A., & Jan-Ola, Ö. (2007). *Socio-onomastics the pragmatics of names*. Amsterdam, NL: John Benjamins.

WorldBank. (2021, 10/25/2021). *São Tomé and Príncipe profile*. Retrieved from https://www.worldbank.org/en/country/saotome/overview