

Commentary

# Ideophones and Realia in a Santome/Portuguese Bilingual Dictionary

Gabriel Antunes de Araujo <sup>1,2</sup> 

<sup>1</sup> Department of Portuguese, University of Macau, Macau, China; gabrielaraujo@um.edu.mo

<sup>2</sup> Departamento de Letras Clássicas e Vernáculas, Universidade de São Paulo, São Paulo 05508-220, Brazil

Received: 27 August 2020; Accepted: 1 November 2020; Published: 10 November 2020



**Abstract:** In this work, we discuss how Araujo & Hagemeyer's Santome/Portuguese bilingual dictionary defines and describes ideophones and realia lemmata. We show that ideophones were listed individually along with their expression counterparts. Realia lemmata (words and expressions for culture-specific items) or specialized lexical units were presented in their Santomean forms, followed by a description of their endemic specificities. Many realia items from Santome can also be found in Portuguese. We conclude that the authors contribute to the lexicographic record of ideophones, lexical items that did not exist in Portuguese, but relevant to the language and culture of Santome. On the other hand, with the documentation of realia entries, they collaborate for the validation of lexical units (originating in Santome) in the local vernacular variety of São Tomé and Príncipe's Portuguese, a common historical practice in Portuguese lexicography.

**Keywords:** ideophones; realia; Santome; Portuguese; lexicography

## 1. Introduction

Santome, or *Forro* (cri<sup>1</sup>), is a Portuguese-based creole language spoken by about 35,000 people in the Democratic Republic of São Tomé and Príncipe in the Gulf of Guinea (INE 2012). The goal of this article is to lexicographically discuss the definition and description of two linguistic units in Araujo and Hagemeyer's (2013) bilingual<sup>2</sup> Santome/Portuguese dictionary: the ideophone and realia lemmata. We will show that the documentation of these lexical units in the Santome/Portuguese dictionary may help the description of the lexicon of some vernacular varieties of São Tomé and Príncipe Portuguese.

Four Portuguese-based creole languages emerged in the Gulf of Guinea in the sixteenth century: Santome and Angolar on the island of São Tomé, Lung'le on Príncipe, and Fa d'Ambô on Annobón (Ferraz 1979; Hagemeyer 2009; Bandeira 2017). All four languages are descendants of the Portuguese-based Gulf of Guinea Proto-Creole, which was created at the beginning of the sixteenth century during contact between Portuguese colonists and the African populations brought as slaves to the island of São Tomé. Isolation, migration of certain groups from the island, and linguistic contributions from the African languages by way of the constant renewal of the enslaved population contributed to the proto-creole's speciation. Santome developed in the colonial centers on the island of São Tomé, while Angolar (aoa) was the language of the descendants of runaway slaves who escaped from the plantations and founded maroon communities. Proto-Creole speakers were taken to the

<sup>1</sup> ISO 639-3: 2007 provides a three-letters code to identify a language. Therefore, when a language here is mentioned for the first time, it will be identified by its unique three-letters code. See: *Codes for the representation of names of languages—Part 3: Alpha-3 code for comprehensive coverage of language* (<https://www.iso.org/standard/39534.html>) and Eberhard et al. (2019).

<sup>2</sup> Here, a bilingual dictionary is understood as 'a type of dictionary which relates the vocabularies of two languages together by means of translation equivalents' (Hartmann and James 2001, s.v. *bilingual dictionary*).

islands of Príncipe and Annobón, where local conditions contributed to diversification and gave rise to Lung'le (pri) and Fa d'Ambô (fab), respectively (Bandeira et al. 2019).

In recent years, these languages have undergone grammatization processes<sup>3</sup> thanks to the production of dictionaries such as the *Dicionário livre do santome-português* (Araujo and Hagemeyer 2013). Santome and its sister languages (Angolar and Lung'le) have autonomous linguistic structures and are mutually unintelligible, even though they are languages with a Portuguese lexical base and share a common origin. However, all these languages have ideophones<sup>4</sup> (Bartens 2000), and due to the nature of their ecolinguistic systems, they also have realia items or specialized lexical units words and expressions for culture-specific items related to endemic fauna and flora, culture, and technology (Vlahov and Florin 1969, p. 432; Bergenholtz and Tarp 1995, p. 15). In general, understanding the linguistic nature of ideophones and the realia items in any minority language allows the lexicographer to address these phenomena in a way that encompasses linguistic facts, respecting the characteristics of working languages and the scientific accuracy of the dictionary.

The Santome-Portuguese bilingual dictionary's audience is formed of readers who are monolingual in Portuguese, bilingual Portuguese/Santome, or even scholars. Furthermore, Portuguese has been the official language of São Tomé and Príncipe since 1975 and is presently spoken by more than 98% of the population (Araujo 2020). Although it is the smallest of all Portuguese-speaking countries, this language is currently spoken by the majority of its 200,000 inhabitants. It has been reported that 80% of people under 20 years of age may speak only Portuguese (Bouchard 2019; Araujo 2020, p. 193).

In most Western languages with an alphabetic tradition, the development of dictionaries in print, unlike digital media (software and applications), has promoted the use of access structure based on a systematic order. For this reason, classifying each lemma is crucial in assisting a user who is looking for lexical items in printed lexicographical works. In digital media, automated search and cross-referencing systems make this issue less salient. There is little discussion of complex units and encyclopedic lemmata in Portuguese dictionaries in the literature (Bacelar do Nascimento 2006, 2013; Martins 2013; Bacelar do Nascimento et al. 2006). However, in the history of lexicography, the recording of complex lexical items in both monolingual and bilingual dictionaries underwent distinct phases. The items in Portuguese lexicography are treated as main lemmata and sometimes as subentries, confined to encyclopedic or specialized dictionaries or simply ignored. Bluteau's (1712–1728) *Vocabulário português e latino, áulico, anatômico, architectonico, etc.*, for example, realia lemmata in Brazilian Portuguese without equivalents in European Portuguese classified as 'Brazilian words' (*brasileirismos*), creating a new and prolific category (see Frankenberg-Garcia 2017).

The current text is organized as follows. In Section 1, we define ideophones, exemplify their usage in Santome, and discuss the solutions for presenting such lexical units in a bilingual dictionary. In Section 2, we approach the realia lemmata in Santome and describe how they were documented. The final section presents concluding remarks.

## 2. Ideophones

### 2.1. Definition

The term ideophone has been used in the literature to define a lexical unit with a high degree of syntactic rigidity formed by a noun and a qualitative, a verb and a predicate, or an adverb commonly related to colors, sounds, smells, actions, states, or intensity (Araujo 2009; Bartens 2000; Costa 2017; Doke 1935; Voeltz 1971; Westermann 1907). Bartens (2000, p. 14) argued that ideophones

<sup>3</sup> Auroux (1992) grammatization "is the process by which all the fluxes and flows [flux] through which symbolic (that is, also, existential) acts are linked, can be discretized, formalized and reproduced. The most well-known of this process is the writing of a language" (Stiegler 2011, p. 172).

<sup>4</sup> The term 'ideophone' here differs from *sound symbolism*, as used in Bantu linguistics. See Bartens (2000) for a comprehensive analysis of ideophones in Creole languages.

typically present sounds and combinations of sounds not found in the phonological inventory of the language. This may be the case with the ideophones analyzed here because they may not respect some phonotactic characteristics of their languages. Costa (2017, p. 8), for example, mentions some special suprasegmental features such as vowel lengthening in Santome. Nevertheless, ideophones in Santome have been translated into Portuguese in the literature (Araujo 2009; Araujo and Hagemeijer 2013; Costa 2017; Ferraz 1979) using superlatives or the formula ‘very x’ or ‘x-ish,’ where ‘x’ is the first lexical item of the ideophone lexical unit. In Portuguese or English translations, adverbs such as ‘genuinely’, ‘strongly’, and ‘extremely’ or lexical items indicating intensity or quality of excess or repetition have been traditionally used.

In addition, even though there are ideophones whose origin may be onomatopoeic, the etymological origin of this phenomenon may be related to simple lexical items in African languages (Ferraz 1979; Costa 2017; Hagemeijer and Ogie 2011). However, due to the historical process, speakers do not have any intuition about the historical link, and the original terms themselves have been changed through morpho-phonological reinterpretation or loss and addition of segmental and suprasegmental material (Ferraz 1979). Finally, ideophones should always occur in expressions or in connection with their nouns or verbs. Therefore, speakers do not utter ideophones without a context or without anaphoric syntactic relations in the discourse (Araujo 2009; Costa 2017). An ideophone and its related lexical unit are called an ‘expression,’ and the second part itself is the ideophone, as proposed by Araujo and Hagemeijer (hereinafter, A & H 2013 in the examples). In (1), we present examples of isolated ideophones in Santome and examples of their full expression are shown in (2).<sup>5</sup>

- (1) **lululu** [lulu'lu] (id.) Cf. **pletu lululu**. (A & H 2013, s. v.).  
**kla** ['kla] (id.) Cf. **kota kla**. (A & H 2013, s. v.).  
 (2) **pletu lululu**<sup>6</sup> ['pletu lulu'lu] (expr.) 1. Pretíssimo. 2. Nigérrimo. (A & H 2013, s. v.).  
**kota kla**<sup>7</sup> [kɔ'ta 'kla] (expr.) Cortar ao meio. (A & H 2013, s. v.).

Furthermore, ideophones occur in settings restricted to only one lexical item, such as in (3), and rarely form units with more than one word.<sup>8</sup> When this occurs, the lexical items related to the ideophone are cognates or semantically related, as in (4).<sup>9</sup>

- (3) **moli mogogogo**<sup>10</sup> ['moli mɔ'gɔmɔ'gɔ] (expr.) Molíssimo. (A & H 2013, s. v.).  
 (4) a. **blanku fenene**<sup>11</sup> ['blɛku fɛnɛ'nɛ] (expr.) Branquíssimo. (A & H 2013, s. v.).  
 b. **klalu fenene**<sup>12</sup> ['klalu fɛnɛ'nɛ] (expr.) Claríssimo. (A & H 2013, s. v.).

Some authors (Brindle 2011; Friesen 2016; Voeltz 1971; Yakpo 2019) have stressed the onomatopoeic character of ideophones, but in Santome, there are linguistic motivations other than the mimetic character of a sound. Nevertheless, some ideophones, though not all, may be related to onomatopoeia:

<sup>5</sup> All examples are from Araujo and Hagemeijer (2013). In the text, they appear exactly as in the dictionary. English simplified equivalents are presented in the footnotes.

<sup>6</sup> **pletu lululu** ‘very black’. See **pletu** ‘black’.

<sup>7</sup> **kota kla** ‘to cut in half’. See **kota** ‘to cut’.

<sup>8</sup> The present work does not aim to provide an in-depth discussion of ideophones. For this, see Bartens (2000) for ideophones in several languages. For ideophones in Santome, see Araujo (2009) and Costa (2017). All examples were translated from Portuguese.

<sup>9</sup> Although rare, it is not impossible for an ideophone to be used with semantically unrelated lexical units, as can be seen in **kulu dii** ‘very dark, deep night’ and **da son din** ‘falling flat’ (Costa 2017, p. 51).

**moli mogogogo** ‘very soft’. See **moli** ‘soft’.

<sup>11</sup> **blanku fenene** ‘very white, whitish’. See **blanku** ‘white’.

<sup>12</sup> **klalu fenene** ‘very clear’. See **klalu** ‘clear’.

- (5) **flêbê blublublu**<sup>13</sup> [fle'be blublu'blu] (expr.) Entrar em ebulição. (A & H 2013, s. v.).  
**tlêmê tatata**<sup>14</sup> [tle'me tata'ta] (expr.) Tiritar. (A & H 2013, s. v.).

## 2.2. Ideophones in Santome

Ferraz (1979, pp. 75–78) was the first to address ideophones in Santome. Although the author does not effectively offer a comprehensive description, Ferraz stated that the ideophone is an element of a category that groups any word for which the modifying form of a verb or noun is repeated or duplicated (Araujo 2009, p. 25). In addition, Ferraz (1979) listed ideophones as reduplicated words that do not behave as expressions, such as **leve-leve** ‘more or less’. However, in general, Ferraz’s observation about partial or full reduplication of syllables is generally attested, with few exceptions, as shown by the following Santome examples.

- (6) a. **blanku fenene** ‘very white’ > **blanku** ‘white’  
 b. **pletu lululu** ‘very black’ > **pletu** ‘black’  
 c. **kentxi zuzuzu** ‘very hot’ > **kentxi** ‘hot’  
 d. **ximpli talitali** ‘tasteless’ > **ximpli** ‘simple’

Araujo (2009) argued that not every ideophone in Santome contains a reduplicated form because there are monosyllabic ideophones (7a), ideophones in which all syllables are different (7b), with all syllables repeated whenever the syllables repeat (7c), with only the two initial syllables repeated (7d), with only the two final syllables repeated (7e), and a monosyllabic word with a long vowel (7f). Costa (2017, p. 68) stated that 54% of ideophones contain repeated or reduplicated parts.

- (7) a. **kota kla** [kɔ'ta 'kla] (expr.) Cortar ao meio. (A & H 2013, s. v.).  
 b. **seku klakata**<sup>15</sup> ['seku klaka'ta] (expr.) Sequíssimo. (A & H 2013, s. v.).  
 c. **lêdê zazaza**<sup>16</sup> [le'de zaza'za] (expr.) 1. Ardor de uma ferida. 2. Ardor provocado por pimenta. (A & H 2013, s. v.).  
 d. **pya babaka**<sup>17</sup> ['pya baba'ka] (expr.) Estar pasmo. (A & H 2013, s. v.).  
 e. **betu blalala**<sup>18</sup> [betu blala'la] (expr.) Escancarado. (A & H 2013, s. v.).  
 f. **kulu dīīī**<sup>19</sup> ['kulu 'dīīī] (expr.) Ecuríssimo. (A & H 2013, s. v.).

In the next section, we will address how Araujo and Hagemeyer’s work included lexical lemmata with ideophones in their dictionary.

## 2.3. Ideophones in Araujo and Hagemeyer’s Work

Araujo and Hagemeyer (2013)’s a lexicographic work that consists of a Santome-Portuguese bilingual dictionary and a Portuguese–Santome list. In this dictionary, the lemmata, arranged in alphabetical order, contain the lemma (presented in the official orthography for this language) followed by the phonetic transcription, word class, and equivalents. Some lemmata, such as the functional lexicon, are accompanied by examples. Lemmata with ideophones and dialectal variation forms are

<sup>13</sup> **flêbê blublublu** ‘to boil’. See **flêbê** ‘to boil/to simmer’.

<sup>14</sup> **tlêmê tatata** ‘to shiver’. See **tlêmê** ‘to tremble’.

<sup>15</sup> **seku klakata** ‘very dry, completely dry’. See **seku** ‘dry’.

<sup>16</sup> **lêdê zazaza** ‘burning wound, spiciness of a chili pepper’. See **lêdê** ‘to burn’.

<sup>17</sup> **pya babaka** ‘to be mesmerized’. See **pya** ‘to see’.

<sup>18</sup> **betu blalala** ‘wide open’. See **betu** ‘open’.

<sup>19</sup> **kulu dīīī** ['kulu 'dīīī] ‘very dark’. See **kulu** ‘dark’.

connected to full expressions, including the ideophone itself and a basic form, respectively. The lemmata on fauna and flora contain their scientific names wherever possible. In the Portuguese–Santome reverse list, words and phrases in Portuguese only refer to their equivalent in Santome without a word class, phonetic form, etc.

The listing of any lexical item in a dictionary requires a reflection on its nature, which includes not only its meaning or equivalents in the case of bilingual dictionaries but also its form and function. Consequently, ideophones pose a challenge for lexicographers because they are part of lexical units with a compulsory association to another name. Therefore, they are not free forms carrying meaning but a combined form. Thus, in the dictionary, ideophones must be listed twice: individually and alongside their lexical counterpart. In the first case, ignoring the meaning of the ideophone or its lexical counterpart, the individual lemma of an ideophone would allow the user to look for it (knowing it or not) in alphabetical order and relate it to its counterpart. If the ideophone is not listed individually, an unfamiliar user who does not know it or ignores its full expression could not perform the alphabetical search. Additionally, the inclusion of the multilexical lemma would allow the user to know the ideophone's unique or limited association pattern. Araujo and Hagemer (2013) opted to list ideophones individually, as in (8), referring the user to the compound unit, as in (8).

- (8) a. **bligidi** [bliɡi'di] (id.) Cf. **ba bligidi**. (A & H 2013, s. v.).  
 b. **sonosono** [sɔ'nɔsɔ'nɔ] (id.) Cf. **liku sonosono**. (A & H 2013, s. v.).  
 c. **ba bligidi**<sup>20</sup> ['ba bliɡi'di] (expr.) 1. Desmoronar. 2. Despencar. (A & H 2013, s. v.).  
 d. **liku sonosono**<sup>21</sup> ['liku sɔ'nɔsɔ'nɔ] (expr.) Riquíssimo. (A & H 2013, s. v.).

In (8), the structure of the lemmata contains the lemma (in bold), a phonetic transcription, and the abbreviation of the word class (*id.*) for the ideophone. Subsequently, there is the abbreviation Cf. (compare or see), followed by the complete multilexical unit containing the related item and its ideophone. Therefore, the isolated ideophone presents a cross-reference to the full expression. In turn, the data in (8) contains the lemma of the complete multilexical lemma and the phonetic transcription. Additionally, the word class is (*expr.*), that is, an expression, followed by its equivalents.

Many African languages have ideophones, and they are represented in lexicography in different ways. For example, in Brindle's (2011, 2017) dictionary and grammatical outline of Chakali (cli), a Gur language spoken in Ghana, it is stated that the majority of ideophones in that language function like qualifiers, intensifiers, or adjunct adverbials. However, Brindle does not describe ideophones as members of multilexical units. In general, all ideophones are listed as basic lemmata with a full description and an example, as in (9). In Brindle's English–Chakali dictionary, shown in (10), a simpler co-reference list is provided.

- (9) **felfel** [félfél] *ideo.* manner of movement, as a lightweight entity, applicable to leaves, animals and humans •òtʃòjɛ̃ kaalì félfél. She ran away, lightly. (Brindle 2017: s.v.).

**ganagana** [gàràgàrà] *ideo.* to be under-ripe and still stiff and hard • títòtì à mójògò nòjò há dò gànàgàrà. Do not pluck the mango fruit, it is still stiff. (Brindle 2017: s.v.).

- (10) **motion** (manner) **felfel** *ideo.* (Brindle 2017: s. v.).

**ripe** (under-) **ganagana** *ideo.* (Brindle 2017: s. v.).

<sup>20</sup> **ba bligidi** 1. 'to collapse'. 2. 'to plummet'.

<sup>21</sup> **liku sonosono** 'very rich'.

Additionally, Friesen (2016) presented a description of Moloko (mlw), a Chadic language spoken in Cameroon, with an English–Moloko and a Moloko–English Lexicon. Ideophones in Moloko are described, following Doke (1935, p. 188), as sound symbolism, where they “evoke the ‘idea’ of a sensation or sensory perception (action, movement, color, sound, smell, or shape). As such they are often onomatopoeic” (Friesen 2016, p. 115). In the Moloko–English Lexicon, as in (11), ideophones are described with sensory images.

(11) **abəlgamay** *id.* n. the way a sick person walks. (Friesen 2016, p. 405).

**bakaka** *id.* spicy hot taste. (Friesen 2016, p. 406).

**gəraw** *id.* idea of cutting something through the middle. (Friesen 2016, p. 411).

In the English–Moloko Lexicon, ideophones are listed as ‘ideas of’ something (12).

(12) **idea of the way a sick person walks** *abəlgamay*.

**idea of cutting something through the middle** *gəraw*. (Friesen 2016, p. 432).

**idea of spicy hot taste** *bakaka*. (Friesen 2016, p. 432).

Pichi (fpe), an English-based creole spoken in the Republic of Equatorial Guinea, is another African language with ideophones, as described by Yakpo (2009, 2019). The author claims that, in Pichi, ideophones “are words with expressive semantics and particular structural characteristics” (Yakpo 2019, p. 443). Furthermore, Yakpo adds: “It is therefore difficult to ascertain how widespread the use of these ideophones is, and whether some of them are sound symbolic ad hoc creations, whether they are carried over from other languages used by the speaker, or whether they form part of the lexicon of Pichi” (Yakpo 2019, p. 443). Even though ideophones are documented in the Pichi–English list in Yakpo’s work, as in (13), many are not present in the English–Pichi word list, such as **bwa** and **gbin**. When ideophones are listed, they have a simple co-reference item. In this sense, the search for an ideophone in Yakpo’s printed word lists may not be an easy task.

(13) **bwa** *ideo.* sound of gushing water. (Yakpo 2019, p. 562).

**gbin** *ideo.* sound of a hard and sudden blow. (Yakpo 2019, p. 566).

(14) a. **katakátá** *ideo.* (be) (hyper-)active, hectic.

**katakátá mán** hyper-active, hectic man. (Yakpo 2019, p. 567).

b. **active (hyper~)** *ideo.* **katakátá**. (Yakpo 2019, p. 579).

Considering that the grammatical category of ideophones does not exist in the Portuguese language, the solution presented by Araujo and Hagemer (2013), as in example (8), is convenient for allowing users to find both the ideophone alone and the complete unit in the minority language. Therefore, it is a useful solution, especially to readers of Portuguese or other languages with no ideophones. However, isolated ideophones are not listed in the Portuguese/Santome list because they do not belong to the Portuguese local vernacular lexicon. Nonetheless, names and their ideophones are listed with their full Portuguese meanings, as in (15).

(15) **branquíssimo**<sup>22</sup> Blanku fenene. (A & H 2013, s. v.).

**cortar ao meio**<sup>23</sup> Kota kla. (A & H 2013, s. v.).

<sup>22</sup> Very white.

<sup>23</sup> To cut into halves.



### 3. Realia

The formation of the Portuguese-based Proto-Creole of the Gulf of Guinea and its later speciation into four languages (Santome, Lung'le, Angolar, and Fa d'Ambô) in the sixteenth century are related to the colonial system that was implemented in the then uninhabited islands of São Tomé and Príncipe and Annobón.<sup>24</sup> The thousands of African slaves kidnapped from the mainland, mainly from the regions of the Niger Delta, Congo, and Angola, and their coexistence with Portuguese settlers of European origin promoted the emergence of the proto-creole. However, the dozens of mother tongues of the slaves—including their unique ways of naming the world—as well as the Portuguese language of the European settlers and their cultures and the linguistic agency of members of those communities—associated with the very nature of the new island's environments and their endemic fauna and flora—promoted the naming of a new world without a parallel in the Portuguese language. Thus, like any human language, Santome has lexical items that reflect their unique specificities, known as *realia* or 'cultureme' (Vlahov and Florin 1969; Xatara and Seco 2014) or defined as specialized lexical units (Bergenholtz and Tarp 1995).

Therefore, Araujo and Hagemeijer (2013) applied a solution for *realia* lemmata that correlates a lexical item (simple or compound) to an identical word in local Portuguese or to a full description when the *realia* in Santome does not circulate in the vernacular Portuguese. Thus, their dictionary also documents lexical items in the minority language related to the vernacular Portuguese of São Tomé and Príncipe. In fact, most *realia* items from Santome already circulate in the local Portuguese varieties. In (16), for example, Araujo and Hagemeijer (2013, p. 41) present a lexical lemma naming a species of pepper endemic to the region of West Africa without an equivalent in Portuguese. Consequently, the dictionary chose to repeat the name of the plant in italics—thus, establishing a 'Santomean Portuguese word' (a language fact proper to the Portuguese influenced by the Santome language or the ethnic group Santome/*Forro*)—followed by its scientific name (listed whenever possible in Araujo and Hagemeijer's work) written in bold and italics. Thus, in (16a), a name is listed for a fauna item, an endemic tree, as an example of *realia*. If a local fauna or flora item has a Portuguese word, the authors simply documented it, such as (16b), 'pau-sabrina'. However, (16c) and (16d) show an interesting contrast: (16c) is a *realia* item in Santome with the same equivalent in vernacular Portuguese, however, the equivalent of (16d) is a combination of a Santomean-originated *realia* and a Portuguese expression (*do-mato*), creating a new item in Portuguese.

- (16) a. **fya-bôba-pikina** [fja 'boba pi'kina] (n.) *Fya-bôba pikina. Piper capense.* (A & H 2013, s. v.).  
 b. **gofi-d'ôbô** [gofi do'bo] (n.) Pau-sabrina. *Musanga cecropioides.* (A & H 2013, s. v.).  
 c. **gligô** [gli'go] (n.) Gligô. *Morinda lucida.* (A & H 2013, s. v.).  
 d. **gligô-d'ôbô** [gli'go do'bo] (n.) Gligô-do-mato. *Sacosperma paniculatum.* (A & H 2013, s. v.).

Nonetheless, in (17), the name of a traditional therapist dedicated to the sensory examination of urine (through smell, texture, taste, color, and impurities) of patients with kidney diseases or diabetes is a *realia* item. However, differently from a fauna or flora item, it was necessary to propose a definition of this word. In these three cases, the Santome term was repeated in the dictionary because the lexical item is itself the equivalent word in Portuguese.

- (17) **pyadô-zawa**<sup>25</sup> [pja'do 'zawa] (n.) Pyadô-zawa. Terapeuta tradicional que examina a urina. (A & H 2013, s. v.).

Therefore, São Tomé and Príncipe's vernacular varieties of Portuguese have already loaned many words from Santome and other national languages. Moreover, in daily work with the informants,

<sup>24</sup> Territory of the present Republic of Equatorial Guinea.

<sup>25</sup> **pyadô-zawa** 'traditional therapist that examined urine'. See **pyadô** 'observer'.

it was often difficult to separate the items from one another given the increasing use of Portuguese. Users were fully aware of the influence of Portuguese and were therefore sometimes confused about the source of many endemic items. However, working with the community is crucial for the acceptance of the dictionary by the local residents. The materials in this Santome dictionary may thereby feed other lexicographical works focusing on African countries that use the Portuguese language (see [Bacelar do Nascimento 2013](#)) with unique items from the linguistic reality of São Tomé and Príncipe. The documentation of a ‘Santomean Portuguese word’ (a language fact proper to the Portuguese influenced by the Santome language or its ethnic group) has a parallel in the Luso-Brazilian lexicography tradition, just as ‘Brazilian Portuguese words’ have been populating the Portuguese language ([Bluteau 1712–1728](#); [Frankenberg-Garcia 2017](#)).

Additionally, the Santome/Portuguese dictionary contains lexical items that refer to festivals, religious rituals, flora, fauna, technologies, etc., which require a description in the dictionary because they are unique to São Tomé and Príncipe’s linguistic and cultural environment. The example in (18) refers to a traditional medicine given to women ready to give birth. Whereas in (19), *sôwô* is the name of a local dish, prepared with breadfruit, cassava, plantain, fish and aromatic herbs, served with manioc flour or baked bananas.

- (18) **magita-pali**<sup>26</sup> [ma'gita pa'li] (n.) *Magita-pali*. Medicamento tradicional para parturientes. (A & H 2013, s. v.)
- (19) **sôwô**<sup>27</sup> ['sowo] (n.) *Sôwô*. Prato típico que pode, alternativamente, ser preparado com fruta-pão, mandioca, inhame, matabala, batata-doce, banana-pão, com peixe e óleo de palma e algumas ervas aromáticas, acompanhado de farinha de mandioca ou banana assada. (A & H 2013, s. v.)

Items (20)–(23), for example, are correlated to one another. *Realia* in (20) describes a spiritual debt (*dêvê*). The example in (21) describes a ritual led by a master (a traditional therapist) who acts out a ceremony resulting in a payment of a spiritual debt through prayers and offerings deposited at crossroads or other sacred sites. Furthermore, *fêgula* in (22) is the name of a human body-like ritual toy used in that ceremony. Note that *fêgula* is the word for ‘figure’ too, as pointed out by its second meaning. Item (23), another *realia*, names the person who has a spiritual debt to be paid (meaning 1) and a performer of this rituals (meaning 2).

- (20) **dêvê**<sup>28</sup> [de've] (n.) Dívida espiritual. (A & H 2013, s. v.)
- (21) **paga-dêvê**<sup>29</sup> [pa'ga de've] (n.) *Paga-dêvê*. Ritual conduzido por um mestre que simula o pagamento de uma dívida, através de orações e oferendas depositadas em encruzilhadas e outros locais. (A & H 2013, s. v.)
- (22) **fêgula**<sup>30</sup> [fe'gula] (n.) 1. Boneco utilizado no ritual do *paga-dêvê*. 2. Figura. (A & H 2013, s. v.)
- (23) **pagadô-dêvê**<sup>31</sup> [paga'do de've] (n.) 1. Devedor do *paga-dêvê*. 2. Quimbandeiro. (A & H 2013, s. v.)

Like any lexicographical work, the dictionary discussed here should be open to review depending on advances in knowledge. In fact, [Araujo and Hagemeyer \(2013\)](#) proposed the lemma shown in (24a) for Santome in 2013. However, on the island of Príncipe, the same species endemic to São Tomé and Príncipe is also called **kobo-peetu** in local Portuguese. Investigations after [Araujo and Hagemeyer’s \(2013\)](#) date of publication have shown that Portuguese settlers had not introduced this species as previously

<sup>26</sup> **magita-pali** ‘a traditional medicine for parturients’.

<sup>27</sup> **sôwô** ‘a traditional dish prepared with bread-fruit, cassava, taro, matabala, sweet potato, plantain, fish, palm oil, and herbs, side dish by cassava flour or baked banana’.

<sup>28</sup> **dêvê** ‘spiritual debt’.

<sup>29</sup> **paga-dêvê** ‘a ritual led by a religious master that simulates the payment of a spiritual debt, through prayers and offers left in crossroads and other places.’

<sup>30</sup> **fêgula** ‘a ceremonial doll used in the **paga-dêvê** ritual’.

<sup>31</sup> **pagadô-dêvê** ‘the person in spiritual debt at the **paga-dêvê** ritual’.



thought, but that it was indeed a new species (Ceriaco et al. 2017). Thus, Agostinho and Araujo's (in press) Lung'Ie/Portuguese bilingual dictionary already includes the new scientific name of that species endemic to the island of Príncipe, improving the previous work<sup>32</sup>.

- (24) a. **koblo-pletu**<sup>33</sup> ['kɔblɔ'plɛtu] (n.) Cobra preta. *Naja melanoleuca*. (A & H 2013, s. v.).  
 b. **kobo-peetu** ['kɔbɔ'pɛ:tu] (n.) 1. Cobra preta. 2. *Kobo-peetu*. *Naja (Boulengerina) peroescobari*. (Agostinho and Araujo n.d., in press).

Therefore, all realia items in the Santome/Portuguese dictionary may be a source for lexicographers who prepare bilingual and monolingual dictionaries in Portuguese.

#### 4. Final Remarks

In this text, based on the case study of one bilingual dictionary of the Portuguese-based creole language Santome, we presented how the authors dealt with the documentation of ideophones and realia items. The task of documenting Santome, a minority and endangered language spoken in São Tome and Príncipe, is urgent. Thus, Santome can benefit from the production of bilingual dictionaries, just as any threatened minority language can (Auroux 1992; Ogilvie 2010). We demonstrated that ideophones can be documented in isolation and with reference to the other members of their multilexical unit. Thus, the user is allowed to search for the ideophone per se and its full expression in the dictionary. Realia lemmata, in turn, are necessary to describe the endemic characteristics of local lexical items. They simultaneously allow the documentation of lexical units in Portuguese and are able to be classified in the future as Santomean Portuguese words.

The publication of bilingual dictionaries of threatened languages is a step towards the documentation and grammatization of minority languages. Moreover, it enables the lexicographic and scientific discussion of phenomena that do not exist or were not explored in Portuguese, especially in the less known vernacular varieties.

**Funding:** This research was funded by University of Macau, grant number SRG-2019-00153-FAH, and University of São Paulo.

**Acknowledgments:** The author thanks the editors and two anonymous reviewers for comments that helped improve this paper.

**Conflicts of Interest:** The author declares no conflict of interest.

#### References

- Agostinho, Ana Livia, and Gabriel Antunes de Araujo. n.d. *Dicionário Lung'Ie/Português*. São Paulo: Humanitas, in press.
- Araujo, Gabriel Antunes de. 2009. Ideofones na língua sãotomense. *Papia* 19: 23–37.
- Araujo, Gabriel Antunes de. 2020. Portuguese Language Expansion in São Tomé and Príncipe: An Overview. *Diadorim* 21: 57–78. [CrossRef]
- Araujo, Gabriel Antunes de, and Tjerk Hagemeijer. 2013. *Dicionário livre do santome-português*. São Paulo: Hedra.
- Auroux, Sylvain. 1992. *A Revolução Tecnológica da Gramatização*. Campinas: Editora da Unicamp.
- Bacelar do Nascimento, Maria Fernanda. 2006. Comparable Corpora and Lexical Variation in the African Varieties of Portuguese. *Alfa* 50: 189–204.
- Bacelar do Nascimento, Maria Fernanda. 2013. Processos de Lexicalização. In *Gramática do Português*. Edited by Eduardo Raposo, Maria Fernanda Bacelar do Nascimento, Maria Antónia Mota, Luísa Segura and Amália Mendes. Lisboa: Fundação Calouste Gulbenkian, pp. 215–46.

<sup>32</sup> We thank these authors for sharing a version of their unpublished book.

<sup>33</sup> **koblo-pletu** 'forest cobra'.

- Bacelar do Nascimento, Maria Fernanda, Amália Mendes, and Sandra Antunes. 2006. Typologies of MultiWord Expressions Revisited: A Corpus-driven Approach. In *Spoken Language Corpus and Linguistic Informatics*. Edited by Yuji Kawaguchi, Susumu Zaima and Toshihiro Takagaki. Amsterdam: John Benjamins, vol. 5, pp. 227–44.
- Bandeira, Manuele. 2017. *Reconstrução Fonológica e Lexical do Protocrioulo do Golfo da Guiné*. Ph.D. dissertation, Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, Brazil.
- Bandeira, Manuele, Gabriel Antunes de Araujo, and Thomas Finbow. 2019. *The Gulf of Guinea Proto-Creole and its Daughter Languages: From Liquid Consonants to Complex Onsets and Vowel Lengthening*. Manuscript: Universidade de São Paulo.
- Bartens, Angela. 2000. *Ideophones and Sound Symbolism in Atlantic Creoles*. Helsinki: Academia Scientiarum Fennica.
- Bergenholtz, Henning, and Sven Tarp. 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam: John Benjamins.
- Bluteau, Raphael. 1712–1728. *Vocabulário Portuguez e Latino, áulico, Anatômico, Architectonico etc.* Coimbra: Collegio das Artes da Companhia de Jesus. Available online: <http://www.brasiliana.usp.br/en/dicionario/edicao/1> (accessed on 5 May 2019).
- Bouchard, Marie-Eve. 2019. Becoming Monolingual: The Impact of Language Ideologies on the Loss of Multilingualism on São Tomé Island. *Languages* 4: 50. [CrossRef]
- Brindle, Jonathan A. 2011. Aspects of the Chakali Language. Ph.D. thesis, Norges Teknisk-Naturvitenskapelige Universitet (NTNU), Trondheim, Norway.
- Brindle, Jonathan. 2017. *A Dictionary and Grammatical Outline of Chakali*. Berlin: Language Science Press. [CrossRef]
- Cerriaco, Luis M. P., Mariana P. Marques, Andreas Schmitz, and Aaron M. Bauer. 2017. The “Cobra-preta” of São Tomé Island, Gulf of Guinea, Is a New Species of *Naja Laurenti* 1768 (Squamata: Elapidae). *Zootaxa* 4324: 21–41. [CrossRef]
- Costa, Patrícia Pardal da. 2017. *Ideofones em Santome*. Master’s dissertation, Universidade de Lisboa, Lisboa, Portugal.
- Doke, Clement Martyn. 1935. *Bantu Linguistic Terminology*. London: Longmans/Green.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2019. *Ethnologue: Languages of the World*, 22nd ed. Dallas: SIL International.
- Ferraz, Luiz Ivens. 1979. *The Creole of São Tomé*. Johannesburg: Witwatersrand University Press.
- Frankenberg-Garcia, Ana. 2017. The Lexicography of Portuguese. In *International Handbook of Modern Lexis and Lexicography*. Edited by Patrick Hanks and Gilles-Maurice De Schryver. Berlin/Heidelberg: Springer, pp. 1–11. [CrossRef]
- Friesen, Dianne. 2016. *A Grammar of Moloko*. Berlin: Language Science Press.
- Hagemeijer, Tjerk. 2009. As línguas de São Tomé e Príncipe. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* 1: 1–29.
- Hagemeijer, Tjerk, and Ota Ogie. 2011. Èdó influence on Santome. In *Creoles, their Substrates, and Language Typology*. Edited by Claire Lefebvre. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 37–60.
- Hartmann, Reinhard Rudolf Karl, and Gregory James. 2001. *Dictionary of lexicography*. London: Routledge.
- INE. 2012. Instituto Nacional de Estatística (INE): São Tomé e Príncipe em Números. São Tomé. Available online: <http://www.ine.st/2012.html> (accessed on 18 May 2019).
- Martins, Vicente de Paula da Silv. 2013. *Estratégias de compreensão de expressões idiomáticas por não nativos do português brasileiro*. Ph.D. dissertation, Universidade Federal do Ceará, Fortaleza, Brazil.
- Ogilvie, Sarah. 2010. Lexicography and Endangered Languages: What Can Europe Learn from the Rest of the World? In *XIV EURALEX Proceedings of the International Congress*. Edited by Anne Dykstra and Tanneke Schoonheim. Leeuwarden and Ljouwert: Fryske Akademy-Afûk, pp. 27–46.
- Stiegler, Bernard. 2011. *The Decadence of Industrial Democracies. Disbelief and Discredit*. Translated by Daniel Ross, and Suzanne Arnold. Malden: Polity Press, vol. 1.
- Vlahov, Sergej, and Sider Florin. 1969. Neperovodimoe v perevode. Realii. *Masterstvo Perevoda* 6: 432–56.
- Voeltz, Erhard Friedrich K. 1971. Toward the Syntax of Ideophone in Zulu. In *Papers in African Linguistics*. Edited by C. Kim and H. Staklke. Edmonton: Champaign, pp. 141–52.
- Westermann, Diedrich. 1907. *Grammatik der Ewe Sprache*. Berlin: D. Reimer.
- Xatara, Claudia, and Mariele Seco. 2014. Culturemas em contraste: Idiomatismos do português brasileiro e europeu. *Revista Domínios de Linguagem* 8: 502–19. [CrossRef]

Yakpo, Kofi. 2009. *A Grammar of Pichi*. Ph.D. dissertation, Radboud University Nijmegen, Nijmegen, The Netherlands. Available online: <http://hdl.handle.net/2066/79407> (accessed on 6 February 2019).

Yakpo, Kofi. 2019. *A Grammar of Pichi*. Berlin: Language Science Press. [CrossRef]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).