

Ideophones and realia in a Santome/Portuguese bilingual dictionary

Ideofones e realia em um dicionário bilíngue Santome/Português

DOI: <https://doi.org/10.24206/lh.v6i3.33700>

Gabriel Antunes de Araujo

Associate Professor, Department of Portuguese, University of Macau, China (where he acts as a Post-graduate Programmes coordinator), and Departamento de Letras Clássicas e Vernáculas, Universidade de São Paulo, Brazil.

E-mail: gabrielaraujo@um.edu.mo

ORCID: <https://orcid.org/0000-0001-7337-3391>

ABSTRACT

In this work, we discuss how Araujo & Hagemeyer's Santome/Portuguese bilingual dictionary (ARAUJO; HAGEMEIJER, 2013) defines and describes ideophones and *realia* entries. We show that ideophones were listed individually along with their collocation counterparts. *Realia* entries (words and expressions for culture-specific items) 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 (originated 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.

RESUMO

Neste texto, discutimos como o dicionário bilíngue Santome/Português (ARAUJO; HAGEMEIJER, 2013) define e representa os ideofones e as entradas *realia*. Mostraremos que os ideofones foram listados, no dicionário, individualmente e conjuntamente com suas colocações. Já as entradas *realia* requereram a documentação de itens lexicais próprios da língua santome e da sua cultura, muitas delas presentes também no português, seguidas de definições de suas especificidades endêmicas. Concluímos que os autores contribuem com o registro lexicográfico dos ideofones, itens lexicais inexistentes no português, mas importantes na língua santome e na sua cultura. De outro, com a documentação das entradas *realia*, colaboram com o registro de unidades lexicais da variedade vernácula local da língua portuguesa próprias de São Tomé e Príncipe, oriundas do santome, prática histórica comum na lexicografia portuguesa.

Palavras-chave: Ideofones. *Realia*. Santome. Português. Lexicografia.

Introduction

Santome, or *Forro* (CRI¹), 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 Hagemeijer's (2013) bilingual² Santome/Portuguese dictionary: the ideophone and *realia* entries. We will show that the documentation of these lexical units in 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 Ano Bom (FERRAZ, 1979; HAGEMEIJER, 2009). 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 islands of Príncipe and Ano Bom, where local conditions contributed to diversification and gave rise to Lung'le (PRI) and Fa d'Ambô (FAB), respectively (BANDEIRA; ARAUJO: FINBOW, 2019).

In recent years, these languages have undergone grammatization processes³ thanks to the production of dictionaries such as the *Dicionário livre do santome-português* (ARAUJO; HAGEMEIJER, 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 ideophones⁴ (BARTENS, 2000), and due to the nature of their ecolinguistic systems, they also have *realia* items,

¹ ISO code 369-3 is a standardized way to represent the names of languages (see EBERHARD; SIMONS; FENNIG,

² 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; JAMES, 2001, *s.v. bilingual dictionary*).

³ Auroux's (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).

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

words and expressions for culture-specific items related to endemic fauna and flora, culture, and technology (VLAHOV; FLORIN, 1969, p. 432). 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, p. 192). Although it is the smallest of all Portuguese-speaking countries, this language is currently spoken by the majority (circa 98%) of its 200,000 inhabitants. Besides, 80% of people under 20 years of age may speak only Portuguese (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 entries in Portuguese dictionaries in the literature (BACELAR DO NASCIMENTO et al., 2006, 2013; MARTINS, 2013; MENDES 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 entries and sometimes as subentries, confined to encyclopedic or specialized dictionaries or simply ignored. Bluteau's (1712-1728) *Vocabulário portuguez e latino, áulico, anatômico, architectonico, etc.*, for example, classified *realia* items in Brazilian Portuguese without equivalents in European Portuguese 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* entries in Santome and describe how they were documented. The final section presents concluding remarks.

1. Ideophones

1.1 *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 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; HAGMEMEIJER, 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 related to onomatopoeia, the etymological origin of this phenomenon may be related to simple lexical items in African languages (COSTA, 2017; FERRAZ, 1979; HAGEMEIJER; 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 collocations 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 Hagemeyer (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).⁵

- (1) **lululu** [lulu'lu] (id.) Cf. **pletu lululu**. (A & H, 2013, *s. v.*).
kla ['kla] (id.) Cf. **kota kla**. (A & H, 2013, *s. v.*).

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

- (2) **pletu lululu**⁶ [ˈpletu luluˈlu] (expr.) 1. Pretíssimo. 2. Nigérrimo. (A & H, 2013, s. v).
kota kla⁷ [kɔˈta ˈkla] (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.⁸ When this occurs, the lexical items related to the ideophone are cognates or semantically related, as in (4).⁹

- (3) **moli mogogogo**¹⁰ [ˈmɔli mɔˈgɔmɔˈgɔ] (expr.) Molíssimo. (A & H, 2013, s. v).
 (4) a. **blanku fenene**¹¹ [ˈblɛ̃ku fɛnɛˈnɛ] (expr.) Branquíssimo. (A & H, 2013, s. v).
 b. **klalu fenene**¹² [ˈklalu fɛnɛˈnɛ] (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:

- (5) **flêbê blublublu**¹³ [flɛˈbɛ blublɔˈblu] (expr.) Entrar em ebulição. (A & H, 2013, s. v).
tlêmê tatata¹⁴ [tlɛˈmɛ tataˈta] (expr.) Tiritar. (A & H, 2013, s. v).

1.2 Ideophones in Santome

Ferraz (1979, p. 75–78) was the first author to address ideophones in Santome. Although he 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

⁶ **pletu lululu** ‘very black.’ See **pletu** ‘black.’

⁷ **kota kla** ‘to cut in half.’ See **kota** ‘to cut.’

⁸ 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.

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

¹⁰ **moli mogogogo** ‘very soft.’ See **moli** ‘soft.’

¹¹ **blanku fenene** ‘very white, whitish.’ See **blanku** ‘white.’

¹² **klalu fenene** ‘very clear.’ See **klalu** ‘clear.’

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

¹⁴ **tlêmê tatata** ‘to shiver.’ See **tlêmê** ‘to tremble.’

duplicated (ARAUJO, 2009, p. 25). In addition, Ferraz (1979) listed ideophones as reduplicated words that do not behave as collocations, 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 (7), ideophones in which all syllables are different (7), with all syllables repeated whenever the syllables repeat (7), with only the two initial syllables repeated (7), with only the two final syllables repeated (7), and a monosyllabic word with a long vowel (7). Costa (2017, p. 68) stated that 54% of ideophones contain repeated or reduplicated parts.

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

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

1.3 Ideophones in Araujo and Hagemeyer (2013)

Araujo and Hagemeyer’s a lexicographic work in which the vehicle-language is Portuguese. It consists of a Santome-Portuguese bilingual dictionary and a Portuguese-Santome reverse list. In this

¹⁵ **seku klakata** ‘very dry, completely dry.’ See **seku** ‘dry.’

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

¹⁷ **pya babaka** ‘to be mesmerized.’ See **pya** ‘to see.’

¹⁸ **betu blalala** ‘wide open.’ See **betu** ‘open.’

¹⁹ **kulu dīīī** ['kulu 'dīīī] ‘very dark.’ See **kulu** ‘dark.’

dictionary, the entries, 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 entries, such as the functional lexicon, are accompanied by examples. Entries with ideophones and dialectal variation forms are connected to full expressions, including the ideophone itself and a basic form, respectively. The entries 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. As a result, 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 entry 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 entry would allow the user to know the ideophone's unique or limited association pattern. Araujo and Hagemeyer (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**²⁰ [ba bliɡi'di] (expr.) 1. Desmoronar. 2. Despençar. (A & H, 2013, *s. v.*)
 d. **liku sonosono**²¹ [liku sɔ'nɔsɔ'nɔ] (expr.) Riquíssimo. (A & H, 2013, *s. v.*)

In (8), the structure of the entries 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 entry and the phonetic transcription. Additionally, the word class is (*expr.*), that is, an expression, followed by its equivalents.

²⁰ **ba bligidi** 1. 'to collapse.' 2. 'to plummet.'

²¹ **liku sonosono** 'very rich.'

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 entries with a full description and an example, as in (9). In Brindle's English-Chakali, 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 • ò tʃóje ò kààli félfél. She ran away, lightly. (BRINDLE, 2017, *s. v.*)
ganagana [gàràgàrà] *ideo.* to be under-ripe and still stiff and hard • tí tóti à móngò nójì ò há dó gànàgànà. 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.*)

Additionally, Friesen (2016) presented a description of Moloko (MLW), a Chadic language spoken in Cameroon, with an English-Moloko and a Moloko-English glossary. 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 glossary, 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 glossary, 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 Hagemeyer (2013), as in (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**²² Blanku fenene. (A & H, 2013, *s. v.*).
cortar ao meio²³ Kota kla. (A & H, 2013, *s. v.*).

2. Realia

The formation of the Portuguese-based Proto-Creole of the Gulf of Guinea and its later speciation into four languages (Santome, Lung’Ie, Angolar, and Fa d’Ambô) in the sixteenth century

²² Very white.

²³ To cut into halves.

are related to the colonial system that was implanted in the then uninhabited islands of São Tomé and Príncipe and Ano Bom.²⁴ 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’ (NADAL, 2009; VLAHOV; FLORIN, 1969; XATARA; SECO, 2014). These lexical items are a challenge for lexicographers in general (PINHEIRO, 2018).

Therefore, Araujo and Hagemeyer (2013) applied a solution for *realia* entries 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 Hagemeyer (2013, p. 41) present a lexical entry naming a species of pepper endemic to the region of West Africa without an equivalent in Portuguese. Consequently, the lexicographers 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 Hagemeyer’s work) written in bold and italics. Thus, in (16), 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 (16), ‘pau-sabrina.’ However, (16) and (16) show an interesting contrast: (16) is a *realia* item in Santome with the same equivalent in vernacular Portuguese, however, (16) is a combination of a Santomean-originated *realia* and a Portuguese word, 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).

²⁴ Territory of the present Republic of Equatorial Guinea.

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**²⁵ [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, it was often difficult to separate the items from one another given the increasing use of Portuguese. As such, users were fully aware of the influence of Portuguese and were sometimes confused about the source of many endemic items. Thus, 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 interested in 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. Example in (19) refers to a traditional medicine given to women ready to give birth. While in (18), **sôwô** is the name of a local dish, prepared with breadfruit, cassava, plantain, fish, aromatic herbs, served with manioc flour or baked bananas.

(18) **magita-pali** [ma'gita pa'li] (n.) *Magita-pali*. Medicamento tradicional para parturientes. (A & H, 2013, s. v.)

²⁵ pyadô-zawa 'traditional therapist that examined urine.' See pyadô 'observer.'

- (19) **sôwô** [ˈsowo] (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** [de've] (n.) Dívida espiritual. (A & H, 2013, s. v.)
- (21) **paga-dêvê** [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** [fê'gula] (n.) 1. Boneco utilizado no ritual do paga-dêvê. 2. Figura. (A & H, 2013, s. v.)
- (23) **pagadô-dêvê** [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. As a matter of fact, Araujo and Hagemeyer (2013) proposed the entry shown in (24) 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 thought but that it is indeed a new species (CERÍACO et al., 2017). Thus, Agostinho and Araujo's (in press) Lung'le/Portuguese bilingual dictionary already includes the new scientific name of that species endemic to the island of Príncipe, improving the previous work.

- (24) a. **koblo-pletu** [ˈkɔblɔ ˈpletu] (n.) Cobra preta. *Naja melanoleuca*. (A & H, 2013, *s. v.*).
 b. **kobo-peetu** [ˈkɔbɔ ˈpɛtu] (n.) 1. Cobra preta. 2. *Kobo-peetu. Naja (Boulengerina) peroescobari*. (Agostinho and Araujo, 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.

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* entries, 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.

Acknowledgments

This study was supported by a CNPq grant (310463/2016-5), University of São Paulo, and University of Macau (SRG-2019-00153-FAH). Two anonymous reviewers provided very useful suggestions and comments. Following their suggestions, the manuscript was thoroughly improved.

References

- AGOSTINHO, A. L.; ARAUJO, G. A. **Dicionário Lung'le/Português**. São Paulo: Humanitas, in press.
- ARAUJO, G. A. Ideofones na língua sãotomense. **Papia**, v. 19, p. 23–37, 2009.
- ARAUJO, G. A.; HAGEMEIJER, T. **Dicionário livre do santome-português**. São Paulo: Hedra, 2013.
- AUROUX, S. **A revolução tecnológica da gramatização**. Campinas: Editora da Unicamp, 1992.
- BACELAR DO NASCIMENTO, M. F. Comparable Corpora and Lexical Variation in the African Varieties of Portuguese. **Alfa**, v. 50.2, p. 189–204, 2006.
- BACELAR DO NASCIMENTO, M. F. Processos de Lexicalização. *In*: RAPOSO, E.; BACELAR DO NASCIMENTO, M. F.; MOTA, M. A.; LUÍSA SEGURA, A. M. (eds). **Gramática do Português**. Lisboa: Fundação Calouste Gulbenkian, 2013. p. 215–246.
- BACELAR DO NASCIMENTO, M. F.; MENDES, A.; ANTUNES, S. Typologies of MultiWord Expressions Revisited: A Corpus-driven Approach. *In*: KAWAGUCHI, Y.; ZAIMA, S.; TAKAGAKI, T. (eds). **Spoken Language Corpus and Linguistic Informatics**. Amsterdam: John Benjamins, v. 5, 2006. p. 227–244.
- BANDEIRA, M.; ARAUJO, G. A.; FINBOW, T. **The Gulf of Guinea Proto-Creole and its Daughter Languages: from Liquid Consonants to Complex Onsets and Vowel Lengthening**. Manuscrito: Universidade de São Paulo, 2019.
- BARTENS, A. **Ideophones and Sound Symbolism in Atlantic Creoles**. Helsinki: Academia Scientiarum Fennica, 2000.
- BLUTEAU, R. **Vocabulário portuguez e latino, áulico, anatômico, architectonico etc.** Coimbra: Collegio das Artes da Companhia de Jesus, 1712–1728.
<http://www.brasiliana.usp.br/en/dicionario/edicao/1>. Accessed: 05 May, 2019.
- BRINDLE, J. A. **Aspects of the Chakali language**. PhD thesis, Norges Teknisk-Naturvitenskapelige Universitet (NTNU), 2011.
- BRINDLE, J. A. **A dictionary and grammatical outline of Chakali**. Berlin: Language Science Press, 2017. DOI: 10.5281/zenodo.344813
- CERÍACO, L. M. P.; MARQUES, M. P.; SCHMITZ, A.; BAUER, A. M. The “Cobra-preta” of São Tomé Island, Gulf of Guinea, Is a New Species of *Naja Laurenti* 1768 (Squamata: Elapidae). **Zootaxa**, v. 4324.1, p. 21–41, 2017. <https://doi.org/10.11646/zootaxa.4324.1.7>.
- COSTA, P. P. **Ideofones em Santome**. Dissertação de mestrado, Universidade de Lisboa, 2017.
- DOKE, C. M. **Bantu Linguistic Terminology**. London: Longmans/Green, 1935.

- EBERHARD, D. M.; SIMONS, G. F.; FENNIG, C. D. (eds.). **Ethnologue: Languages of the World**. Twenty-second edition. Dallas, Texas: SIL International, 2019.
- FERRAZ, L. I. **The Creole of São Tomé**. Johannesburg: Witwatersrand University Press, 1979.
- FRANKENBERG-GARCIA, A. The Lexicography of Portuguese. In: HANKS, P.; DE SCHRYVER, G.-M. (eds). **International Handbook of Modern Lexis and Lexicography**. Berlin/Heidelberg: Springer, 2017. p. 1-11. https://doi.org/10.1007/978-3-642-45369-4_66-1.
- FRIESEN, D. **A grammar of Moloko**. Berlin: Language Science Press, 2016.
- HAGEMEIJER, T. As línguas de São Tomé e Príncipe. **Revista de Crioulos de Base Lexical Portuguesa e Espanhola**, v.1, p. 1-29, 2009.
- HARTMAN, R. R. K.; JAMES, G. **Dictionary of lexicography**. London: Routledge, 2001.
- INE. 2012. Instituto Nacional de Estatística (INE): São Tomé e Príncipe em Números. São Tomé: 2001. <http://http://www.ine.st/2012.html>. Accessed: 18 May 2019.
- MARTINS, V. P. S. **Estratégias de compreensão de expressões idiomáticas por não nativos do português brasileiro**. Tese de doutorado, Universidade Federal do Ceará, 2013.
- NADAL, L. L. Los culturemas: ¿unidades lingüísticas, ideológicas o culturales? **Language Design**, v. 11, p. 93-120, 2009.
- OGILVIE, S. Lexicography and Endangered Languages: What Can Europe Learn from the Rest of the World?. In: DYKSTRA, A.; SCHOONHEIM, T. (eds). **XIV EURALEX Proceedings of the International Congress**. Leeuwarden/Ljouwert: Fryske Akademy-Afûk, 2010. p. 27-46.
- PINHEIRO, M. G. **Descrição e análise dos culturemas do português brasileiro**. Dissertação de mestrado. UFC, 2018.
- STIEGLER, B. **The Decadence of Industrial Democracies. Disbelief and Discredit**, v. 1. Translated by D. ROSS and S. ARNOLD. Malden: Polity Press, 2011.
- VLAKHOV, S.; FLORIN, S. Neperovodimoe v perevode. *Realii*. **Masterstvo perevoda**, v. 6, 1969, Moskva: Sovetskij pisatel (1970), p. 432-456. Print.
- VOELTZ, E. Toward the Syntax of Ideophone: Zulu. In: KIM, C.; STAKLKE, H. (eds). **Papers in African Linguistics**. Edmonton: Champaign, 1971. p. 141-152.
- XATARA, C. M.; SECO, M. Culturemas em contraste: idiomatismos do português brasileiro e europeu. **Revista Domínios de Lingu@gem**, v. 8.1, p. 502-519, 2014. <https://doi.org/10.14393/DL15-v8n1a2014-29>.
- YAKPO, K. **A grammar of Pichi**. PhD dissertation, Radboud University Nijmegen, 2009. <http://hdl.handle.net/2066/79407>. Accessed: 06 Feb 2019.
- YAKPO, K. **A Grammar of Pichi**. Berlin: Language Science Press, 2019. DOI: 10.5281/zenodo.2546450

WESTERMANN, D. **Grammatik der Ewe Sprache**. Berlin: D. Reimer, 1907.