Linguistic Typology at the Crossroads 4-1 (2024): 381-444 ISSN 2785-0943

Ideophones in Upper Guinea Creoles: a comparative study

NICOLAS QUINT[1], NOËL BERNARD BIAGUI[2]
[1]CNRS, UMR 8135 LANGAGE, LANGUES ET CULTURES D'AFRIQUE
[2]CENTRE DE LINGUISTIQUE APPLIQUEE DE DAKAR (UCAD)

Submitted: 02/04/2023 Revised version: 01/03/2024
Accepted: 04/03/2024 Published: 12/06/2024

Abstract

The Upper Guinea Creoles (UGCs) are a family of closely related Afro-Portuguese languages, comprising three branches: continental (Casamance and Guinea-Bissau), insular (Cape Verde) and ABC (Dutch Antilles). Several continental and insular UGC varieties make use of a specific set of adverbs which can be called "ideophones" following Dingemanse's (2012) definition of the term. This paper aims at providing a comprehensive study of UGC ideophones. Based on a database collected from native speakers, it characterizes the main phonological, morphosyntactic and semantic features of ideophonic items in UGCs. In addition, it investigates the origins of UGC ideophones and compares the use and behavior of this word class in continental and insular UGCs, showing the degree of both African and Portuguese influence on each UGC variety considered.

Keywords: Afro-Portuguese, ideophones, language contact, Niger-Congo, Upper Guinea Creoles.

1. Introduction

Like many Niger-Congo languages, all African members of the Upper Guinea Creoles (UGCs) family, a group of Afro-Portuguese languages spoken in West Africa and in three Dutch Caribbean islands, make use of a specific set of adverbs which can be called "ideophones" (IDEO) following Dingemanse's (2012) definition of the term. Examples (1)-(3), taken from Casamance Creole (CC), illustrate one particular

subcategory of ideophones, namely verb intensifiers, which are highly specialized and each intensify only one or a few verb(s) of the language in question:[1]

(1) CC *braŋku*[2] 'be white' + *fandaŋ* (IDEO) > *braŋku* ***fandaŋ*** 'be **very** white, be **as** white **as snow**'

(2) CC *pretu* 'be black' + *nok* (IDEO) > *pretu* ***nok*** 'be **very** black, be **pitch**-black, be **as** black **as coal**'

(3) CC *kalá* 'be/keep silent' + *mik* (IDEO) > *kalá* ***mik*** 'be **very** silent, be **as** quiet **as a mouse**, be **dead** silent'

The goal of this study is threefold: first, it aims at giving a detailed description of the main phonological, morphosyntactic and semantic characteristics of UGC ideophones; second, it takes stock of the question of their origin; and third, it provides a comparative perspective across two of the three branches of UGCs: the continental branch (Casamance and Guinea-Bissau varieties) and the insular branch (Capeverdean varieties).

The structure of the paper is as follows: in Section 2, we provide the reader with an introduction to the UGC family. Then, in Section 3, we give a brief overview of UGC ideophones, against the background of the typological literature. We explain why our corpus is skewed in favor of verb-intensifying ideophones and mention relevant publications on UGC ideophones. Section 4 provides details about the data and our methodology. Section 5 deals with continental UGC ideophones, whereas Section 6 examines insular UGC ideophones and compares them with their continental counterparts. Section 7 summarizes the main results of this study.

[1] For the intensification of non-verbal elements, see §5.2.2.

[2] Unless otherwise specified, UGC data are transcribed orthographically with phonologically-based conventions that generally follow the rules used by Biagui in his grammar (2018: 31-41) for continental UGC data (Casamance and Guinea-Bissau) and the prescriptions of the ALUPEC (*Alfabeto Unificado para a Escrita do Caboverdiano* = Unified Alphabet for Capeverdean writing) for Capeverdean data. Data drawn from other authors were adapted to these rules. Note that in ALUPEC, the digraphs {dj}, {nh} and {tx} code for one phoneme each, namely /dʒ/~/ɟ/, /ɲ/ and /tʃ/~/c/, respectively.

2. Upper Guinea Creoles: some basic facts

Apart from the specific references mentioned throughout this section, most of the information presented hereafter is based on Biagui et al. (forthcoming), Quint & Moreira Tavares (2019), Jacobs (2012), and Quint (2000b).

2.1. Main subgroupings and varieties

The Upper Guinea Creoles are a group of phylogenetically related Afro-Portuguese languages traditionally spoken in West Africa and in the Dutch Antilles. They comprise three main branches:

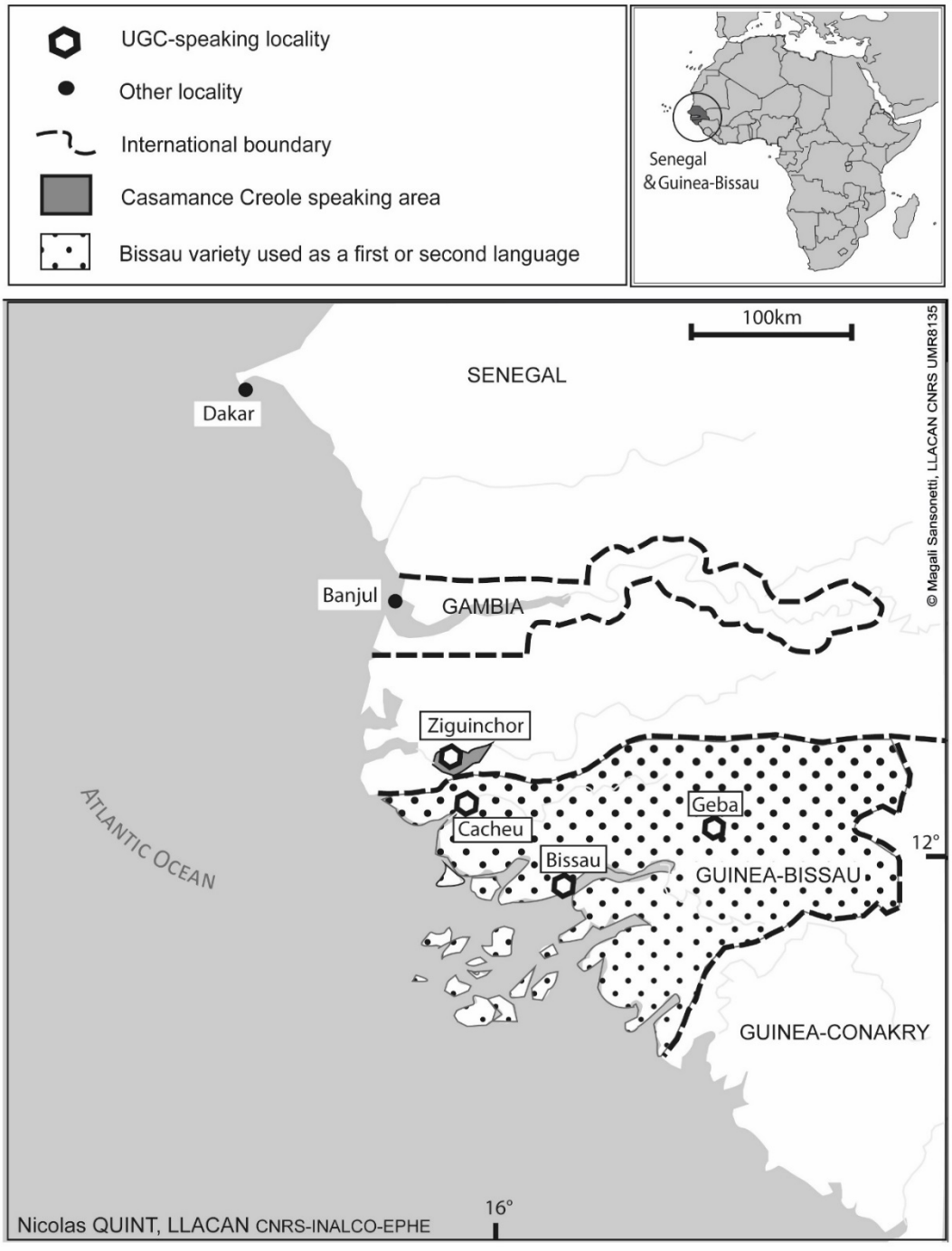
(i) Continental UGCs (pov), spoken in Guinea-Bissau and Senegal (Map 1): in Senegal, only one UGC variety is spoken today, namely Casamance Creole, traditionally used in the city of Ziguinchor and its surroundings (in the region of Lower Casamance). It has approximately 20.000 speakers, 10.000 of whom are native.

In Guinea-Bissau, at least three distinct UGC varieties are recorded: the Bissau variety, originally spoken in the capital city of Guinea-Bissau, has now become the main vehicular language of the whole country. It has nearly 2.000.000 speakers, over 500.000 of whom are native. The Cacheu and Geba varieties, spoken in the eponymous cities situated in Northwestern and Eastern Guinea-Bissau, respectively, are currently on the verge of extinction (as they are being replaced by the Bissau variety), and each of them is used or remembered by only a handful of elderly people (Quint 2023: 452). The Cacheu variety has been shown to be historically related to Casamance Creole (Quint 2023: 451, 453; Quint & Moreira Tavares 2019: 135-139; Biagui 2018: 18) whereas the Geba variety clusters with the Bissau variety.

(ii) Insular (or Capeverdean) UGCs (Map 2) are spoken by 1.000.000 native speakers, 500.000 of whom live in the archipelago of Cape Verde and the rest in diasporic communities, mainly in the US, Portugal, France, Netherlands and Senegal (Quint 2009b). Insular UGCs can be further split into two sub-branches: Sotavento ('Leeward') Capeverdean is spoken in the four southern islands of Brava, Fogo, Maio, and Santiago. Barlavento ('Windward') Capeverdean is spoken in the five northern islands of Boa Vista, Sal, Sant'Antão, São Nicolau, and São Vicente.

(iii) Papiamentu (or ABC UGC, (pap)) is spoken in the three Caribbean islands of Aruba, Bonaire and Curaçao (the ABC islands) by at least 350.000 people including a

diaspora of several tens of thousands of people in the Netherlands. Each island has its own variety.



Map 1: Continental Upper Guinea Creoles.

Map 2: Insular (Capeverdean) Upper Guinea Creoles.

The degree of mutual understanding between the different UGCs is variable: speakers of any two continental varieties can easily chat with each other without any previous contact, and the same is true for speakers of the various ABC varieties. The insular branch shows a higher degree of internal differentiation, in particular between Barlavento and Sotavento: users of rural varieties of each of these subbranches may find it difficult to fully understand each other. Between speakers of different UGC branches, a basic understanding is always possible due to the structural and lexical features shared by all UGCs. However, the linguistic distance between each of the three branches is such that speakers must usually actively learn some elements of each other's varieties before being able to communicate effectively. In other words,

the UGCs can be considered as a group of closely related languages, and each branch of the family as a dialect cluster.

2.2. African and Portuguese origins

All UGCs are characterized by a common Afro-Portuguese lexical core, including a Portuguese component and an African substrate.

Except for contemporary Papiamentu, over 80% of the core lexicon (e.g. Swadesh list) of all UGCs is of Portuguese origin, and therefore UGCs can be described as Portuguese-based Creole languages. Note that this Portuguese component is not derived from today's Portuguese but from classical Portuguese, i.e. the stage of the language that was spoken between 1450 and 1550, at the time when UGCs first developped (see below).

The African substrate is, essentially, derived from the three following languages, in decreasing order of importance: Mandinka (mnk), Wolof (wof/wol) and Temne (tem) (Biagui et al. forthcoming; Quint & Moreira Tavares 2019: 124-126; Rougé 1999). These three languages all belong to the Niger-Congo phylum. Wolof and Temne belong to the Atlantic branch of this phylum while Mandinka belongs to the Mande branch.

It is the existence of this Afro-Portuguese lexical core, combined with a series of striking phonological and structural similarities, which has led scholars to posit the existence of proto-UGC, an ancestor common to all UGCs, which was different from Portuguese.[3] Proto-UGC must have developed during the second half of the 15th century and was in all likelihood fully shaped from 1550 onwards. At the beginning, proto-UGC may have been used as a pidgin (i.e. a non-native contact variety) allowing exchanges between Portuguese sailors and West African people. At any rate, it soon became the native language of some Afro-Portuguese communities, intially in Cape Verde, on the island of Santiago, the first Capeverdean island to have been settled by the Portuguese (ca. 1460) and therefore the most plausible cradle for all UGCs. From then on, UGC varieties developed on the island of Fogo (settled between 1480 and 1490, see Moreira Tavares (2020: 24-25)) and then progressively on the remaining islands of the archipelago of Cape Verde.

Continental varieties must have developed in the second half of the 16th century (Biagui et al. forthcoming), from an offshoot of the incipient Capeverdean (insular) UGC. As all continental UGCs are obviously closely related and share features not

[3] For a first temptative reconstruction of proto-UGC lexicon and grammar, see Quint (2000b: 197-208, 307-318).

found in other UGCs, it is reasonable to posit the existence of a common ancestor for these varieties, namely proto-continental UGC, which must have appeared in Cacheu, the first Portuguese permanent settlement on West African shores.

Papiamentu (or ABC) UGC emerged in the first half of the 17th century, when the Dutch colonized Curaçao and then the two other islands (Aruba and Bonaire) where it is still spoken.

UGCs are mostly analytic languages, and make a much more limited use of bound morphology than Portuguese (and more generally Romance languages). They all follow a strict SVO, determiner + noun, noun + genitive word order.

3. Ideophones in UGCs: a brief overview

3.1. Defining the category of ideophones in UGCs

In UGCs, as in many languages of the world, ideophones are a subtype of adverbs that typically combine with a verb and depict different semantic nuances associated with this verb, such as intensification (1-3), immediacy (4) or different sensory modalities (see Dingemanse 2012: 663 for more details), such as sounds (in which case the ideophones may also be called "onomatopoeias"[4]) (5) or visual patterns (6).

(4) Cacheu (Guinea-Bissau) Creole
(a) *labantá* 'stand up' + *fakat* (IDEO) > *labantá* ***fakat*** 'stand up **at once**'
(b) *sai* 'get out' + *fut* (IDEO) > *sai* ***fut*** 'get out **at once/quickly**'

(5) Casamance Creole (CC)
(a) *kay* 'fall' + *furbap* (IDEO) > *kay* ***furbap*** 'fall **producing the typical noise of something sinking into rice bran**'
(b) *ñemé*[5] 'chew' + *cákum-cákum* (IDEO) > *ñemé* ***cákum-cákum*** 'chew **noisily**'

[4] The concepts of sound ideophone and onomatopoeia do not always overlap in all languages. Within the scope of this paper, we will consider that all UGC sound ideophones can also be considered as onomatopoeias, i.e. "words that originate as imitations of sounds" (Dingemanse 2019: 28). This does not necessarily imply that all UGC onomatopoeias also are ideophones.

[5] For verbs such as CC *ñemé* 'chew' (5b) or Santiago Capeverdean *fasi* 'make' (6a) and *odja* 'see' (6b), which have transitive uses, one could also argue that the ideophones fill an object slot and are not adverbs. However, such ideophones do not behave as prototypical nouns, as (i) they cannot combine

(6) Santiago Capeverdean
(a) *fasi* 'make' + *menhi-menhi* (IDEO) > *fasi* ***menhi-menhi*** 'have **interference**, be **jammed** (television)'
(b) *odja* 'see' + *blu* (IDEO) > *odja* ***blu*** 'see **pink elephants**, have **alcohol-related hallucinations**'

UGC ideophones match the typological characterization elaborated by Dingemanse in 2012, who defines ideophones as "marked words that depict sensory imagery" (Dingemanse 2012: 655). He refined the definition of ideophone in 2019, as a "member of an open lexical class of marked words that depict sensory imagery" (Dingemanse 2019: 16). While it is indisputable that ideophones form an open lexical class in many Niger-Congo languages (e.g. in Koalib (kib), see Quint 2018), this property is not necessarily shared by all UGCs, in particular insular UGCs, hence our preference for the 2012 definition.

3.2. Previous studies on UGC ideophones

UGC ideophones have already been studied in some detail for Guinea-Bissau Creole[6] and briefly mentioned for Capeverdean (Quint 2000b: 107-109; 2008: 72-75) and Casamance Creole (Biagui 2018: 265, 372-373; Biagui & Quint 2013: 47; Dalphinis 1986: 107-108). This material mainly consists of lexeme lists with brief comments. Furthermore, only three of the cited authors have published lists of more than 20 ideophones: Kihm (1995) provides 35, Scantamburlo (1999) 51 and Biagui (2018) 62 ideophones. This paper aims at providing a more complete coverage of the category of ideophones across Upper Guinea Creoles.[7]

---

with nominal determiners (such as demonstratives, numerals, adjectives, etc.) and (ii) generally cannot occupy a nominal slot in a sentence (although the specific behavior of *menhi-menhi* is more ambiguous to this respect, criterion (i) is still valid for this item). Hence, in a sequence such as *odja blu* (6b), from a syntactic point of view, *blu* behaves like a non-ideophonic adverb such as 'well' or 'clearly' and not like a noun such as 'car', 'water' or 'elephant'.

[6] See Quint (2023: 460-461), Truppi & Costa (2019), Scantamburlo (1981: 66-67; 1999: 189-191), Couto (1994: 102-104; 1995: 207-215), Childs (1994: 265-266, 277), Kihm (1994: 76-78), Doneux & Rougé (1988: 27-28), Mbodj (1979: 79; 1984: 58) and Wilson (1962: 62).

[7] Papiamentu (i.e. the third branch of UGCs, see §2.1) is not taken into account. Indeed, as far as we can tell, Papiamentu does not seem to have an ideophonic word class comparable to what will be described hereafter.

4. Data provenance and methodology

This study builds on two main datasets, for continental and insular UGCs respectively.

4.1. The continental dataset

For the continental Upper Guinea Creoles dataset, the following varieties are taken into account (Map 1): Bissau, Cacheu and Geba (Guinea-Bissau) and Casamance (Senegal). In practice, the Casamance variety was used as the standard against which other varieties are compared. As Biagui, himself a Casamance Creole native speaker, had already collected 81 ideophonic items in his dictionary (Biagui in preparation) – the longest list so far compiled for any UGC variety – Quint used these items and systematically checked their semantic equivalents for all other continental varieties. For this, several native speakers and available published sources, mainly Scantamburlo[8] (1999), were consulted (Table 1). All CC ideophones and their other continental UGC counterparts are given in Appendix 1 (Tables 1.1 to 1.5).

Variety	Source	Number of consultants[9]	Code
Casamance	Biagui (to appear) + personal knowledge	1	CC
Geba	fieldwork (2019)	1	Geba
Cacheu	fieldwork (2019)	3	Cacheu1/2/3
Bissau	fieldwork (2019)	2	Bissau1/2
	Scantamburlo (1999)	1	Bissau3

Table 1: Origin of the data for each UGC variety.

The main reasons for the choice of Casamance Creole as the point of departure for the study of these four continental varieties were (i) the availability of a dataset compiled earlier by Biagui and (ii) the fact that he could himself re-check this data and provide in-depth insights as to the behavior of continental UGC ideophones.

[8] In two cases at least (*/far/ and */caŋ/, see Tables 1.3 and 1.4 in Appendix 1), we also resorted to Scantamburlo (2002) to check ideophones that were not listed in Scantamburlo (1999).

[9] Regarding Scantamburlo's dictionary, the data obviously come from more than one consultant. However, as it is a single written source and the specific contribution of each consultant is not mentioned, we will make the approximation of considering that this publication represents one dataset, and therefore is the equivalent of one consultant.

The main reason for the choice of the remaining varieties (i.e. Bissau, Cacheu and Geba) was the desire to provide an exhaustive coverage of continental UGCs, as these varieties, together with Casamance Creole, are the only continental varieties dating back to the settlement period that are still spoken today (see also Quint & Moreira Tavares 2019: 118).

4.2. The insular (Capeverdean) dataset

For Capeverdean, we checked two varieties, namely Santiago and Fogo (Map 2), and looked for ideophonic lexemes comparable to the ones found by Biagui in Casamance Creole. For Santiago, we mostly drew on Quint's own lexical databases, whether published (Quint 1998; 1999) or not. A secondary source was Rougé (2004). The Santiago ideophones considered in this study are given in Appendix 2 (Table 2.1). For Fogo, we relied on the material gathered by Moreira Tavares that was published in Quint & Moreira Tavares (2019). Furthermore, for the sake of comparability, the semantic equivalents of all 81 ideophones attested in Casamance Creole were also systematically checked with one Santiago native speaker.

The main reasons for the choice of Santiago and Fogo varieties were (i) availability of data and (ii) the fact that both Santiago and Fogo varieties are Sotavento (southern) Capeverdean, in which the African element is particularly important, when compared with northern Capeverdean (Quint 2000b: 71-97). Assuming that UGC ideophones were probably an African feature (an intuition that was confirmed by the present study, as shown below), it appeared preferable to choose more Africanized Capeverdean varieties.

4.3. The overrepresentation of verb intensifiers

In this study, we have clearly favored a semantic subcategory of ideophones, namely verb intensifiers (see also §3.1). They represent 61 items out of a list of 81 Casamance Creole ideophones prepared by Biagui, i.e. 75% of the total. The main reason underlying this imbalance is the specific semantic relation between a verb-intensifying ideophone and the verb it modifies. A given verb-intensifying ideophone rarely collocates with more than one or two verbs and, conversely, a given verb is rarely intensified by more than one or two different ideophones. Consequently, the semantic relation between a verb and its intensifying ideophone is salient and clearly

perceived by most native speakers of any UGC variety, which also renders verb-intensifying ideophones easier to elicit for field linguists, when compared with other semantic types of ideophones. In our database, we see, for instance, that several sound ideophones can be associated with the action of FALLING, according to the type of noise produced by this action, e.g. *furbap* (in rice bran, see (5a)), *bup* (someone on a hard surface), *tim* (a metallic object). In other words, verb-intensifying ideophones can be considered, at least in UGCs but also in other languages (e.g. in Koalib, see Quint 2018), as a prototypical semantic type of ideophones, which probably accounts for the fact that our comparative database has a strong bias towards verb-intensifying ideophones. As a matter of fact, all previous work devoted to UGC ideophones also mentions a majority of verb-intensifying ideophones in the examples and lists.

Furthermore, when we began this comparative work, we were mostly focused on intensifying ideophones and tended to discard other semantic types from the ideophone category, in particular sound ideophones. For instance, the equivalents of Casamance Creole sound ideophones *furbap* (5a) and *cákum-cákum* (5b), which Biagui (forthcoming) had already recorded in his dictionary of Casamance Creole, were not systematically checked for other varieties and could therefore not be taken into consideration in this comparative study.

5. Ideophones in the continental branch

5.1. Phonology

As observed in many Niger-Congo languages (Creissels 1994: 30, Quint 2006: 101; 2009a: 91; 2018: 181-190), the phonological characteristics of continental UGC ideophones significantly differ from other lexical items of these languages. This difference is conspicuous at least at four levels: prosody, syllabic structure, vowel pattern and consonant inventory.[10]

5.1.1 Prosody

Nearly all continental UGC verb-intensifying ideophones are pronounced with a clearly recognizable high tone which extends across all syllables of the word. This

[10] These four levels are explicitly mentioned by Dingemanse (2012: 656) as typically marked for many-ideophones across the languages of the world.

$(H)_n$ all-high tone profile is most plausibly a Mandinka feature (§5.4). It contrasts with all other lexical items of the language, which follow a stress pattern (inherited from Portuguese). Consider again the verb + ideophone sequence given in (1), here renumbered as (7).

(7) Bissau1 *braŋku* 'be white' + *fandaŋ* (IDEO) > *braŋku* ***fandaŋ*** 'be **very** white, be **as** white **as snow**'

The verb *braŋku* [ˈbraŋku] 'be white' (< Port. *branco* [ˈbrɐ̃ku]), which is a non-ideophonic lexical item, displays an S-U (=stressed–unstressed) stress pattern. The ideophone *fandaŋ* [fándáŋ][11] 'very' (non-Romance), which is an ideophonic lexical item, displays a H-H (high-high) tone pattern.

However, some continental ideophones do have a stress (*not* tonal) prosodic pattern. In (8), in the sequence *ñemé* ***cákum-cákum*** [ɲeˈme ˈcakum ˈcakum] both the verb *ñemé* 'chew' and the sound ideophone *cákum-cákum* are stressed.

(8) CC *ñemé* 'chew' + *cákum-cákum* (IDEO) > *ñemé* ***cákum-cákum*** 'chew **noisily**'

In continental UGCs, most stressed ideophones seem to belong to semantic categories other than verb intensifiers.[12] Considering the overall prosodic system of the language, a stressed ideophone is less marked than a tonal ideophone, which confirms the prototypical character of verb intensifiers by contrast with other types of ideophones.

5.1.2 Syllabic structure

As shown in Table 2, all continental UGC ideophones save one end in a consonant and over two thirds (69%) are monosyllabic.[13] This data is obviously at variance with the well-formedness standard for continental UGC words, where (i) disyllabic items are clearly favoured, as shown by the fact that 65% of Casamance Creole lexical words

[11] In the phonetic transcriptions provided in this paper, an acute accent above a vowel, e.g. [V́], indicates a high tone.

[12] However, there is at least one case of a stressed verb intensifier ideophone in continental creoles: see fn. 32 and 40.

[13] The peculiar syllabic profile of continental UGC ideophones was already noticed by other authors, e.g. Couto (1995: 212), Kihm (1994: 76) and Wilson (1962: 34).

(and 39% of all lexical items[14] of the language) have a C(C)V.C(C)V structure (Biagui 2018: 89), and (ii) C(C)V is the dominant syllable type, with 73% in Casamance Creole (Biagui 2018: 86) and 65,79% in Bissau Creole (Mbodj 1979: 54).

Syllabic structure	Number of ideophones	%	Syllables	Number of ideophones	%
CV	1	1%	1	56	69%
CVC	55	68%			
CV.CVC	15	19%	2	21	26%
CVC.CVC	6	8%			
CV.CV.CVC	2	2%	3	4	5%
CVC.CV.CVC	2	2%			
Total	**81**	**100%**		**81**	**100%**

Table 2: Syllabic structure of continental UGC ideophones (based on Biagui's data for Casamance Creole).[15]

5.1.3 Vowel patterns

Polysyllabic continental UGC ideophones display a special type of vowel harmony: all vowels of a given ideophone are the same, as in (9) and (10). The only exception is seen in (12).

(9) CC *pagá* 'switch off' + *kamaj* (IDEO) > *pagá* [o] ***k*a*m*a*j*** 'switch [sth.] off **suddenly**'

(10) Cacheu1 *beju* 'be old' + *kokorot* (IDEO) > *beju* ***k*o*k*o*r*o*t*** 'be **very** old, be **as** old **as Methuselah**'

5.1.4 Consonant inventory

As a rule, continental Upper Guinea Creoles have no voiced fricatives in their core vocabulary[16] (Quint 2023; Biagui 2018: 55, 79; Kihm 1994: 17-18; Doneux & Rougé

[14] Throughout this paper, we make a distinction between "lexical word", a category including primarily open lexical classes (nouns, verbs…), "morphological (or grammatical) word", a category including closed lexical classes (adpositions, conjunctions…) and "lexical item", a general label including any word of the language, whether "lexical" or "morphological".

[15] The syllabic structure of the other continental UGC varieties is not significantly different.

[16] This trend can also be observed in conservative varieties of Santiago Capeverdean (an insular UGC), see Quint (2000a: 112-114) and §4.1.4.

1988: 6). However, /v/ and /z/ are attested for several varieties in our ideophone data, as shown in Table 3.

	Ideophone					
No.	**CC equivalent**	**Voiced fricative**	**Variety and consultant**	**No. of occurrences**	**Semantic type**	**Associated verb(s)**
(1)	*bap*	*vap*	Bissau1	1	sound	*sintá* 'sit' + *kay* 'fall'
(2)	*kamaj*	*vut*	Bissau1	1	immediate	*pagá* 'switch off'
(3)	*tar*	*vap*	Cacheu3	1	sound/intensifier	*bafatiyá* 'slap'
(4)	*bup*	*vip*	Cacheu1 + Cacheu3	2	sound	*kay* 'fall'
		vup	Geba	1		
(5)	*lip*	*vip*	Geba	1	intensifier	*pisadu* 'be heavy'
(6)	NA[17]	*zip*	Geba	1	intensifier	*gros* 'be fat/thick'
			TOTAL	**8**		

Table 3: Occurrences of continental UGC ideophones containing a voiced fricative.

All in all, /v/ and /z/ appear very rarely in our data (8 occurrences for ca. 500 UGC ideophones, i.e. 1%) and in variants that have very similar phonological shapes, i.e. /vVp/ ~ /vVt/ ~ /zVp/. However, voiced initial fricatives are attested in all continental UGC varieties except Casamance Creole, which, admittedly, is the most basilectal of the group (see Biagui et al. forthcoming; Quint 2023: 451-452, 454, 457). Note also that at least three of the six ideophones for which voiced fricatives are attested are (or can be used as) sound ideophones, which represent only a minority of our sample (§4.3): the sporadic use of voiced initial fricative variants in sound ideophones may be linked with ICONICITY (Dingemanse 2012: 657-660), in this case the necessity to depict as faithfully as possible the sound perceived by the speakers.[18]

5.1.5 Lexical reduplication

Contrary to many other languages (Dingemanse 2012: 656; Quint 2018), in continental UGCs, only a small number of ideophones have a reduplicated base form. In our

[17] Some few ideophones were taken into consideration even though they were missing in our CC list, for more details see §5.3.

[18] Other languages have been reported to show a higher frequency of voiced fricatives in ideophonic items than in the overall lexicon. This is for example the case of /z/ ~ /dz/ (both realizations appearing in free variation) in Pastaza Quichua (qvz) (Nuckolls et al. 2016).

Casamance Creole sample of 25 polysyllabic ideophones (21 disyllabic + 4 trisyllabic, see Table 2), only six ideophones (i.e. approximately a quarter of the total number of polysyllabic ideophones and 6/81 = 7% of all continental ideophones) have a fully (one item, see (11)) or partially (five items, see (10, 12)) reduplicated lexical base form.

(11) CC *termé* 'shake/shiver' + *bok-bok* (IDEO) > *termé* ***bok-bok*** 'shake **like a leaf**'

(12) Geba *jantí* 'hurry up' + *fit-fat* (IDEO) > *jantí* ***fit-fat*** 'hurry up **making great strides**'

5.2. Morphosyntax[19]

5.2.1 Word order

In all continental UGCs, the ideophone always follows the word (generally a verb) whose meaning it modifies (see also Kihm 1994: 76). To the best of our knowledge, no other element can be inserted between an intransitive verb and an ideophone (13)-(14).

(13) CC

gatu	*kay*	***furbap***	*déntur*	*di*	*bidoŋ*
cat	fall.PFV	**IDEO**	inside	of	barrel

'The cat fell **with a soft noise** inside the barrel [full of rice bran].'

(14) CC

kontrá	*ku*	*sol*	*kinti*	*Mariya*	*simí*	***fes***
when	SUB	sun	be.hot.PFV	Mary	disappear.PFV	**IDEO**
niŋgeŋ	*ka*	*torná*	*wojá-l*	*más*		
nobody	NEG	do.again.PFV	see.PFV-O3SG	anymore		

'During the hottest part of the day, Mary disappeared **completely** [and] nobody saw her anymore.'

With transitive verbs, the object is usually inserted between the verb and the ideophone. This object can be either a noun (15) or a pronoun (16)-(17).

[19] In this part, most examples are drawn from Casamance Creole, which Biagui speaks natively (§4.1). Our comparison of continental UGC varieties was mainly limited to the lexicon and, in most cases, the available data does not allow us to say whether the morphosyntactic processes we describe hereafter are limited to CC or extend across all continental UGCs.

(15) Bissau

iransegu *kumé* *karnel* ***buk***
python eat.PFV sheep **IDEO**
'The python swallowed the sheep **whole**.' (Kihm 1994: 76)[20]

(16) CC

i *kumé-l* ***bik***
S3SG eat.PFV-O3SG **IDEO**
'S/he ate it **all.**'

(17) CC

disna *k-e* *jusiyá* *e* *ta* *pasá* *ŋútur* ***maj-maj***
since SUB-S3PL quarrel.PFV S3PL HAB go.by each.other **IDEO.RED**[21]
'Since they quarrelled, they don't talk anymore (lit. since they quarrelled, they go by each other **with a total lack of attention**).'

5.2.2 Modified elements

In an overwhelming majority of cases, the ideophone modifies a verb (see all examples above) or a qualifier (i.e. an adjective behaving as a verb when used in predicative function,[22] see Quint 2023: 457, 459). However, some other parts of speech, such as adverbs (18) and pronouns (19), may also be modified by an ideophone.

(18) CC *didiya* 'at noon' + *lɐ̀m* (IDEO) > *didiya* ***lɐ̀m*** 'at **solar** noon, when the sun is just **at its peak/zenith**'

(19) Cacheu1 *a-mi* 'me' + *kondoŋ* (IDEO) > *a-mi* ***kondoŋ*** '**on** my **own**, being (myself) **hopelessly alone**'

In (19), *kondoŋ* could be itself analysed as a non-ideophonic adverb meaning 'alone', but this same element also combines with verbs (20) in prototypical ideophonic

[20] This usage of *buk* in association with the verb *kumé* 'eat' is not listed in our own database.

[21] For the use of expressive reduplication in Casamance Creole, see §5.2.3.1.

[22] Note that qualifiers can be modified by an ideophone both in predicative and attributive function.

constructions where one can see that the notion of LONELINESS that *kondoŋ* conveys in (19) is linked with the notion of EMPTINESS or LACK OF CONTENT:

(20) Bissau

garafa	*linpu*	***kondoŋ***
bottle	be.clean/empty.PFV	**IDEO**

'The bottle is **totally** empty.' (Kihm 1994: 77)

The existence of (20) justifies the analysis of *kondoŋ* in (19) as an ideophone modifying the pronoun.

5.2.3 Morphology

Although UGC ideophones are clearly adverbs, they can be involved in at least three types of morphological processes: expressive reduplication, vowel lenghthening and ideophone-to-verb derivation.

5.2.3.1 Expressive reduplication

Expressive reduplication is used to strengthen the meaning of a given ideophone (17, 21, 22).

(21) CC *burmeju* 'red' + *wɛk* (IDEO)
> *burmeju* ***wɛk*** '**vivid/bright** red'
> *burmeju* ***wɛk-wɛk*** '**extremely** bright red'

(22) CC susu 'dirty' + *potok* (IDEO)
> *susu* ***potok*** '**very** dirty'
> *susu* ***potok-potok*** '**horribly** dirty'

Ideophones whose lexical form is reduplicated (§5.1.5) can also undergo (partial) expressive reduplication (23).

(23) CC *termé* 'shake/shiver' + *bok-bok* (IDEO)
> *termé* ***bok-bok*** 'shake **like a leaf**'
> *termé* ***bok-bok-bok*** 'shake like a leaf **without stopping**'

5.2.3.2 Final vowel lengthening

Vowel lengthening is another means to increase the expressiveness of an ideophone and it is only attested for trisyllabic ideophones. In all known cases, it is the last vowel that is lengthened (24):

(24) CC *saŋ* 'healthy' + *keŋkereŋ* (IDEO)
> *saŋ keŋkere**ŋ*** '**very** healthy'
> *saŋ keŋkere**eeŋ*** '**extremely** healthy'

5.2.3.3 Ideophone-to-verb derivation

Several Casamance ideophones can be derived into verbs by the addition of the stressed suffix -/ˈi/ (25)-(26).

(25) CC
(a) *pañá* 'thicken' + *tɐ̈kɐ̈p* (IDEO)
> *pañá **tɐ̈kɐ̈p*** 'get **very** thick' (glue, sauce)
> *tɐ̈kɐ̈p**í*** 'get tough'

(b) *e* *añju* ***tɐ̈kɐ̈pí*** *suma* *turu* *nobu*
DEM.PROX new.born.baby **get.tough**.PFV like bull young
'This newborn baby has a body as **tough** as a young bull.'

(26) CC
(a) *termé* 'shake/shiver' + *bok-bok* (IDEO)
> *termé **bok-bok*** 'shake **like a leaf**'
> *bok-bok**í*** 'shake like a leaf'

(b) *bu* *na* ***bok-bokí*** *na* *suma* *miñjer* *bej-a*
S2SG IPFV.PROG **shake** ASS like woman old-F
'You are **shiver**ing like an old woman.'

The suffix -/ˈi/ is very productive in the language and can also be used to derive verbs from other parts of speech, e.g. nouns (27), including borrowings:

(27) CC *fotó* [foˈto] 'photograph' (< French *photo*)
> *fotoí* [fotoˈi] 'take a photo/picture [of s.o./sth.]'

Therefore, the ideophone-to-verb derivation fits into the general morphological framework of Casamance Creole.

5.3. Diatopic variation

5.3.1 Shared ideophonic roots

If we look at the number of shared ideophonic roots between the different UGC varieties (Table 4), taking into account that Casamance Creole was considered as the standard for comparison, we can distinguish four types of distribution profiles, which are discussed hereafter.[23]

Distribution profile	No. of items	Distribution type	No. of items	%
CC-only	22	CC-only	22	26%
CC + 1[24]	4	Minority models	14	17%
CC + 2	5			
CC + 3	5			
CC + 4	11	Common core	39	46%
CC + 5	8			
CC + 6	17			
CC + 7	3			
CC vs. Guinea-Bissau	3	Other distribution	9	11%
no CC vs. Guinea-Bissau	3			
(CC + Guinea-Bissau) vs. Guinea-Bissau	3			
Total	**84**		**84**	**100%**

Table 4: Number of UGC continental varieties sharing the same ideophonic root as Casamance Creole (CC).

In the *'CC-only' type*, 22 CC ideophones (i.e. 26% of the total) have either no semantic counterpart in any other UGC variety or their respective UGC semantic equivalents are not attested in more than one UGC variety each and these UGC semantic

[23] This classification into distribution types is also used in the data appendices.

[24] The digit following 'CC +' indicates the number of consultants from other UGC continental varieties sharing an ideophonic root with Casamance Creole (CC), e.g. 'CC + 1' means that one consultant speaking a continental UGC variety other than CC provided an ideophone whose phonological form is comparable to the CC semantically equivalent ideophone.

equivalents have a lexical root different from CC. The 'CC-only' distribution type underlines the outlier status of Casamance Creole among continental UGCs (Quint & Moreira Tavares 2019: 133-134), probably due to (a) a peculiar Jola and Nyun[25] adstrate and (b) the political boundary, which, for almost 140 years (Biagui & Quint 2013: 41, Biagui 2018: 18; Quint 2023: 451-452, Biagui et al. forthcoming), has been isolating Casamance Creole – spoken in Senegal – from the remaining continental UGCs – all spoken in Guinea-Bissau.

The contrast between CC and the other UGCs is also conspicuous at the phonological level, even when the CC form is obviously related to other UGCs, i.e. for ideophones belonging to the 'common core' type, see entries (30) and (37) in Table 1.3 in the appendix.

(28) Continental UGCs *burmeju* 'red'
> Geba + Cacheu1/2/3 + Bissau1/3 *burmeju* ***wa****k* '**bright** red'
≠ CC *burmeju* ***wɐ****k*

(29) Continental UGCs *moli* 'soft'
> Geba *moli* ***p****ɔcɔk* ~ Cacheu1/2/3 + Bissau1 *moli* ***p****ɔtɔk* ~ Bissau3 *moli* ***p****otok* '**very** soft'
≠ CC *moli* ***b****otok*

In the *'common core' type*, nearly one half (39/81 = 48%) of our standard list of 81 CC ideophones have a close equivalent for at least 4 of the 7 consultants (§4.1) from other UGC varieties. We can therefore reasonably posit that these forms can be traced back at least to proto-continental UGC, i.e. to the second half of the 16th century (§2.2).

In the *'minority model' type*, some CC-related ideophones were found in the data of only a minority (i.e. fewer than four) of consultants from other UGC varieties. In many cases, these common roots are probably also traceable to proto-continental UGC, as it is very plausible that such phonological and semantic coincidences between several UGCs are not due to mere chance. However, a lower number of formal matches implies a higher degree of uncertainty. It is also striking that the 'minority model' type concerns fewer items than the 'common core' and 'CC-only' types, which means that the main split is between ideophones only found in CC and those shared by all

[25] Jola and Nyun are not provided with an ISO Code-639, because there are no satisfactory code for each of these language groups.

(or most) UGCs. The 'minority model' therefore represents an intermediary category, whose existence may be attributable to factors related to the fieldwork setting (e.g. misunderstandings between collector and consultants) or the lower frequency of forms.

The *'other distribution' type* covers three subtypes which deserve to be studied in more detail. In the 'CC vs. Guinea-Bissau' subtype, the Casamance Creole (CC) form contrasts with a form attested in several Guinea-Bissau UGC varieties (30).

(30) Continental UGCs *tesu* 'be solid'
> Geba + Cacheu2 + Bissau1 *tesu* ***kaŋ*** 'be **very** solid' ~ Cacheu3 *tesu* ***taŋ***
≠ CC *tesu* ***liŋ***

In the 'no CC vs. Guinea-Bissau' subtype, a form attested in several Guinea-Bissau UGC varieties is lacking in Casamance Creole (31).

(31) Continental UGCs *negá* 'refuse'
> Geba + Cacheu1/2 + Bissau3 *negá* ***far*** 'refuse **adamantly**'
≠ no known form in CC[26]

Finally, in the '(CC + Guinea-Bissau) vs. Guinea-Bissau' subtype, the CC form clusters with one or several Guinea-Bissau forms and contrasts with the remaining Guinea-Bissau forms (32).

(32) Continental UGCs *sukuru* 'be dark'
> CC + Geba *sukuru* ***mut*** 'be **pitch/completely** dark'
≠ Cacheu3 + Bissau2/3 *sukuru* ***mip***

5.3.2 Variation according to variety

Except for the distinctive character of CC reflected in the 'CC-only' type, which is also linked with the choice of CC as a standard of comparison, there are no significant differences between the other UGC varieties as regards their ideophone inventories.

Bissau2 and Bissau3 are the only two datasets lacking an equivalent for a majority of the 81 CC ideophones (Table 5). However, these results are an experimental

[26] This subtype explains why the total number of ideophones in Table 4 is 84 instead of 81 elsewhere. Here, 84 results from the 81 ideophones listed for CC (§4.1) plus the 3 'no CC vs. Guinea-Bissau' cases.

artifact, because we did not check the complete list of CC ideophones with consultant Bissau2. As a matter of fact, with consultant Bissau2, we only recorded some forms we wanted to check after having worked with Bissau1. The Bissau3 dataset was taken from a publication (Scantamburlo 1999: 189-191, see §4.1), and we could not check missing forms with a consultant.

If we consider the other consultants and varieties, their behavior is quite similar: they lack equivalents for roughly the same proportion of the total list (between 23 to 36 items, i.e. 28% to 44% of the total) and most of them have few or no semantic equivalents for the ideophones belonging to the 'CC-only' type (§5.3.1). Consultant Cacheu3 has the lowest number of missing CC equivalents, which means Cacheu3 has more in common with CC than any other consultant. The specific proximity between CC and Cacheu3 is reflected by several examples, such as (33).

(33) *pisadu* 'be heavy'
> *pisadu* ***sip*** (Geba + Cacheu1/2/3 + Bissau1/2) 'be **as** heavy **as lead**'
≠ *pisadu* ***lip*** [27] (CC + Cacheu3).

One could explain this higher degree of similarity by the fact that the Cacheu and Casamance UGC varieties are historically related (§2.1). Nonetheless, the other available datasets for Cacheu (Cacheu1 and Cacheu2) do not display noticeable differences to Geba and Bissau (or rather Bissau1, the only complete dataset available for Bissau) regarding their number of missing equivalents of CC ideophones. Hence, it is not possible to contrast the ideophone inventories of Cacheu with that of Geba and/or Bissau. UGC ideophones are definitely very similar in all Guinea-Bissau varieties, which is due to their common origin (see §2.2 and §5.3.1) but also to the levelling influence of the Bissau variety, which is about to erase the Cacheu and Geba varieties (§2.1).

[27] Note that *lip* is only attested as an intensifier of *fartá* 'be satisfied' in Cacheu3 and that Cacheu3 also uses *sip* (as other Guinea-Bissau continental varieties). However, given that (i) *lip* is not used by any other Guinea-Bissau variety; (ii) SATISFIED and HEAVY are obviously semantically related; and (iii) the local Cacheu variety is about to be totally replaced by the Bissau variety (§2.1), this specific resemblance between CC and Cacheu3 can be considered as linguistically significant.

	Missing equivalents[28]	
Variety	**Total**	**CC-only**
Geba	30	22
Cacheu1	29	21
Cacheu2	36	18
Cacheu3	23	15
Bissau1	31	18
Bissau2	74	22
Bissau3	47	22

Table 5: Number of missing equivalents of CC ideophones in other UGC continental varieties.

5.4. Origin

Most of the ideophones whose origin is identified come from Mandinka and/or Wolof (see Table 6 and Table 3.1. in Appendix 3).

Continental UGC ideophones		**Ideophones with an attested African etymon**			
Distribution type	**Total**	**Wolof**	**Mandinka**	**Wolof or Mandinka**	**Total**
CC-only	22	**4**	2	1	7
Minority models	14	**5**	3	0	8
Common core	39	2	**10**	5	17
Other distribution	9	**4**	1	2	7
Total	**84**	**15**	**17**	**7**	**39**

Table 6: African-derived continental UGC ideophones according to their distribution type as defined in Table 4.

A significant proportion (39/84 = 46%) of the ideophones studied in this paper have a plausible African etymon. This confirms Couto's (1995: 212) and Kihm's (1994: 76)

[28] We forgot to check the equivalents of some few CC ideophones in Cacheu2 (four items) and Bissau1 (three items), which may account for the fact that these two datasets have a total number of missing equivalents slightly higher than Geba, Cacheu1 and Cacheu3.

intuition that continental UGC ideophones originally come from West African languages.[29]

Wolof (Niger-Congo, Atlantic) and Mandinka (Niger-Congo, Mande) are the only known African sources for continental UGC ideophones. This result is in accordance with the fact that these languages have long been recognized as the main African substrates of UGCs (§2.2). Nevertheless, the fact that Mandinka and Wolof, due to their widespread use as vehicular languages in West Africa, are better described than many other languages of the region – see, in particular, the detailed lexicographic works for Mandinka by Creissels (2012) and Creissels et al. (1982), and for Wolof by Diouf (2003) and Fal et al. (1990) – is also a significant factor accounting for the absolute dominance of these two languages as etymological sources of UGC ideophones.

The ratios of Mandinka- and Wolof-derived continental UGC ideophones are roughly equivalent. This result is at variance with what was found in previous studies that considered UGC African-derived items in all parts of speech (Quint 2000b: 23-34, 110-113; Quint 2008: 32-47; Quint & Moreira Tavares 2019: 124-126). These studies showed a clear dominance of Mandinka-derived items over Wolof-derived items with at least a 2:1 ratio. In other words, continental UGC ideophones seem to be more Wolofized than the other African-derived items of the language. However, if we look at Table 6 in more detail and check the ratio of Mandinka and Wolof for each of the four recognized distribution types, a more nuanced picture emerges.

Mandinka is clearly dominant in the 'common core' type (10 Mandinka-derived vs. 2 Wolof-derived ideophones), i.e. among the ideophones which can be traced back to proto-continental UGC with most certainty. Conversely, Wolof seems to dominate in all other types.

The 'CC-only' type comprises the ideophones considered as typical of Casamance Creole (CC). As CC is the northernmost variety of continental UGC, as the Wolof-speaking area is situated north of the Mandinka speaking area, and as CC is spoken in Senegal where Wolof is the main lingua franca, it comes as no surprise that the Wolof element is stronger in CC than in any other continental UGC.[30]

---

[29] Conversely, it goes against Childs (1994: 265-266), who, after a first exploratory study on Guinea-Bissau Creole, wrote that "[n]one of the [Guinea-Bissau Creole] ideophones were identified in the first languages [=Atlantic and Mande traditional languages] used alongside Guinea-Bissau Creole" and inferred from this observation (which is contradicted by the present data) "that the ideophones were generated internally, rather than borrowed". Wilson (1962: 34) basically upholds Child's position.

[30] For the question of Wolof influence on CC (and more generally on northern continental UGCs), see Biagui et al. (forthcoming) and Quint & Moreira Tavares (2019: 124-126, 135-139).

The 'minority models' assemble ideophones that are less likely to be traced back to proto-continental UGC than those of the 'common core'. The 'other distribution' type is characterized by the fact that not all UGCs share the same lexical root for a given ideophone, which implies that, for most of these cases, at least one of these roots cannot be traced back to proto-continental UGC.

In sum, the Mandinka component is stronger among those ideophones that can be traced back to proto-continental UGC, i.e. the common stage underlying all known continental UGCs. This seems to suggest that the particularly high ratio of Wolof-derived items among continental UGC ideophones is due to the fact that a significant part of these Wolof-derived items made their way into some UGC varieties *after* the period of formation of continental UGCs, i.e. Wolof is both a *substrate* for UGCs – it contributed some ideophonic elements during their period of formation – and an *adstrate*, as it contributed some ideophonic elements after the period of formation. The same applies to Mandinka, which is both a substrate for all UGCs and an adstrate for several contemporary continental UGCs (Quint & Moreira Tavares 2019: 125-126). However, the results given in Table 6 suggest that Mandinka has been a less powerful adstrate than Wolof.[31]

Another argument in favor of the prevalence of Mandinka as a continental UGC substrate is the fact that in Mandinka – a tonal language – ideophones are generally characterized by the same $(H)_n$ all-high tone profile (Creissels 2013: 323; 2012; Creissels et al. 1982: xvii) as continental UGC ideophones (§5.1.1). As Wolof is a stress language, it is probable that the tonal profile of UGC continental ideophones was inherited from Mandinka.[32]

Besides Mandinka and Wolof, at least one continental UGC ideophone is derived from Portuguese (34).

(34) Continental UGCs *kurpu limpu* 'naked', lit. 'body clean'
> Geba + Cacheu1/3 + Bissau1 *kurpu limpu* ***nu*** '**stark** naked'
= CC *kurpu limpu* ***nuŋ***
with UGC *nu/nuŋ* < Port. *nu* [ˈnu] 'naked'

[31] The relative weakness of Mandinka as an adstrate may also be partly explained by the fact that CC (the most Wolof-influenced continental UGC variety, see Quint (2023: 451-452), Biagui et al. (forthcoming)) was chosen as a standard for this comparative study.

[32] Most notably, [ˈmoku], the only known UGC verb-intensifying ideophone having a clear stress pattern on the continent (see fn. 40 and Table 3.1 in Appendix 3), comes from Wolof.

6. Ideophones in the insular branch (Capeverdean)

In this section, a sample of 34 insular UGC ideophones will be systematically compared with their continental counterparts.

6.1. Phonology

6.1.1 Prosody

Contrary to continental UGCs, insular UGC ideophones are prosodically much more integrated into the general rules applying for the language. All insular ideophones are stressed (like any other lexical word in all insular UGCs), and two stress profiles are allowed, oxytonic (35) and paroxytonic (36).

(35) Santiago *abri* 'open' + *uandaŋ* [wɐ̃ˈdɐ̃] (IDEO) > *abri* ***uandaŋ*** 'open **wide** [a door]'

(36) Santiago *sai* 'get out' + *futi* [ˈfuti] (IDEO) > *sai* ***futi*** 'get out **at once/quickly**'

The paroxytonic profile is clearly dominant, as it represents 24.5[33] items (i.e. 82%) out of a total of 30 polysyllabic ideophones.

6.1.2 Syllabic structure

Contrary to continental UGCs, a majority of insular UGC ideophones (72%) contain only C(C)V open syllables (Table 7), and most of them (63%) are disyllabic (Table 8). Insular UGC ideophones therefore exactly fit into the syllabic canon of the language, which is characterized, at least in the two varieties under study, by a preponderance of C(C)V syllables and a majority of disyllables for stressed items (see Quint 2000a: 35-40 for Santiago and Moreira Tavares 2020: 97-101 for Fogo).

Another peculiarity of insular UGC ideophones is the fact that at least six of them (i.e. roughly one out of six) contain one or two CCV syllables, while these complex onsets are not attested in continental UGC ideophones. Actually, CCV syllables are

[33] The reason why, in this section devoted to the phonology of insular UGC ideophones, we are dealing with half-words is due to the fact that one of these ideophones has at least three different variants (*txulupu* [tʃuˈlupu] ~ *txulufu* [tʃuˈlufu] ~ *txufun* [tʃuˈfũ]) with two different stress and syllable patterns (/CVˈCVCV/ (=*txulupu* + *txulufu*) vs. /CVˈCVC/ (=*txufun*)), see entry (31) in Table 2.1 in the appendix. In consequence, we have considered that each of these two patterns counted as 0.5 item in our word counts.

much more frequent in Capeverdean (12% in texts, see Quint 2000a: 393, 396, 399) than in continental UGC varieties (4,5% in texts, see Biagui 2018: 335, 339, 355), which may account for this difference. Note also that, in all Capeverdean ideophones with CCV syllables, this syllable type always appears in the first syllable of the word (37), although it can be repeated in reduplicated ideophones (38).

(37) Santiago *sakédu* 'stand' + *tran* (IDEO) > *sakédu* ***tran*** 'stand **upright**'

(38) Santiago *gordu* 'fat' + *plátxi-plátxi* (IDEO) > gordu ***plátxi-plátxi*** '**very** fat, **as** fat **as a pig**'

Syllabic structure	**No.**	**CCV. and CV. merged**	**No.**	**%**	**Syllabic type**	
CCV.CCV	1	C(C)V.C(C)V	16	47%	open = C(C)V only	72%
CCV.CV	2					
CV.CV	13					
CCV.CV.CV	1	C(C)V.CV.CV	3.5	10%		
CV.CV.CV	2.5					
CCV.CV.CCV.CV	1	C(C)V.CV.C(C)V.CV	5	15%		
CV.CV.CV.CV	4					
CCVC	1	C(C)VC	4	12%	closed = C(C)VC only	24%
CVC	3					
CVC.CVC	4	CVC.CVC	4	12%		
CV.CVC	0.5	CV.CVC	0.5	1%	mixed	4%
VC.CV	1	VC.CV	1	3%		
Total	**34**	**Total**	**34**	**100%**		**100%**

Table 7: Syllabic structure of insular UGC ideophones (based on Santiago data):[34] syllabic types.

CCV. and CV. merged	**No.**	**%**	**Syllables**	**No.**	**%**
C(C)VC	4	12%	1	4	12%
C(C)V.CV	16	47%	2	20.5	63%
CV.CVC	0.5	1%			
CVC.CVC	4	12%			
VC.CV	1	3%			
C(C)V.CV.CV	3.5	10%	3	3.5	10%
C(C)V.CV.C(C)V.CV	5	15%	4	5	15%
Total	**34**	**100%**		**34**	**100%**

Table 8: Syllabic structure of insular UGC ideophones (based on Santiago data): number of syllables.

[34] As discussed in §6.3, Fogo ideophones are far less numerous than their Santiago counterparts.

CVC monosyllabic ideophones (4/34 items = 12%, including CCVC) and ideophones ending in a consonant (8,5/34 items = 25%) are much rarer in insular than in continental UGCs (68% and 99% respectively). This is largely due to the fact that insular UGCs admit a much lower number of consonants in the coda (Quint 2000a: 33-34; Moreira Tavares 2020: 75-76) than continental UGCs (Biagui 2018: 59-63). This contrast is particularly conspicuous for ideophonic roots shared by both UGC branches (39)-(40).

(39) Cacheu3 (continental)/Santiago (insular) *sai* 'get out'
> Cacheu3 *sai **fut*** [fut] 'get out **at once/quickly**', see (4b)
≠ Santiago *sai **futi*** [ˈfut**i**], see (36)

Fut [fut] has a CVC structure and final [t], allowed by all continental UGCs, whereas ***futi*** [ˈfut**i**] ends in a paragogic [i] (and therefore has a CVCV structure), as final [t] is not allowed by Santiago insular UGC.

(40) CC *kumé* (continental)/Fogo (insular) *kumê* 'eat'
> CC *kumé **fep*** [fep] 'eat [sth.] **thoroughly**, eat **every last crumb** [of sth.]'
≠ Fogo *kumê **fépu*** [ˈfɛp**u**]

Fep [fep] has a CVC structure and final [p], allowed by all continental UGCs, whereas ***fépu*** [ˈfɛp**u**] ends in a paragogic [u] (and therefore has a CVCV structure), as final [p] is not allowed by Fogo insular UGC.[35]

6.1.3 Vowel patterns

A rule very similar to the one described for continental UGCs (§5.1.3) also applies to polysyllabic insular UGC ideophones: all vowels of an ideophonic root are identical (41), except for the post-tonic vowel (42), which is probably paragogic, see (39) and (40), and therefore not part of the root.

(41) Santiago *txera* 'smell' + *fututú* (IDEO) > *txera **fututú*** [f**u**t**u**ˈt**u**] 'have a **strong** smell'

[35] Our relatively reduced sample of insular ideophones seems to suggest that the choice of the paragogic vowel in insular UGCs depends on the preceding consonant (preC). If preC = [t, dʒ~ɟ, tʃ~c] (i.e. a dental or palatal consonant), the paragogic vowel is [i] (38). If preC = [b, f, p, k] (i.e. a labial or velar consonant), the paragogic vowel is [u] (39). The only known exception to these rules is Santiago *saki* [ˈsɐki] ~ ***siki*** [ˈsiki] (IDEO) > *sápa saki* ~ *sápa siki* 'cut **clean**'.

(42) Santiago *labánta* ‘stand up’ + *fakati* (IDEO) > *labánta **fakati*** [f**ɐ**ˈk**ɐ**ti] ‘stand up **at once**’, compare with ***fakat*** in Cacheu (4a)

This rule applies to all five polysyllabic ideophonic roots attested in Santiago Capeverdean, all of which have cognates in continental varieties.

6.1.4 Consonantic inventory

Only one known Santiago ideophone can be optionally pronounced with a voiced fricative (43).

(43) CC *intxi* ‘fill’ + *bipu* ~ *vipu* (IDEO) > *intxi **bipu*** or *intxi **vipu*** ‘fill [sth.] **up to the brim**’

Voiced fricatives are equally rare in Santiago core vocabulary (Quint 2000a: 112-114).

6.1.5 Reduplication

While lexically reduplicated ideophones are rare in continental UGCs, they are much more frequent in Santiago Capeverdean. Out of a sample of 29 polysyllabic items (20.5 disyllables + 3.5 trisyllables + 5 quadrisyllables), nine insular UGC ideophones, i.e. close to one third of the total number of polysyllabic ideophones and 9/34 = 26% of all insular ideophones, have a fully (eight items, see (44)) or partially (one item, see (41)) reduplicated lexical base form. All quadrisyllabic Santiago ideophones, e.g. (44), are in fact reduplicated forms.

(44) Santiago *gordu* ‘fat’ + *bódji-bódji* (IDEO) > *gordu **bódji-bódji*** ‘**very** fat, **as** fat **as a pig**’

6.1.6 Comparison of the phonology of insular and continental UGCs

Table 9 sums up the main commonalities and differences between continental and insular UGC ideophones. There is a clear phonological split between the two UGC branches and, in at least some domains (prosody, preferred syllabic structure, final –C, CCV), a greater Portuguese influence probably explains why insular UGCs differ

from continental UGCs. Indeed, Portuguese, like Capeverdean, is a stress language (prosody), with a dominant disyllabic word-pattern (preferred syllabic structure), few final consonants (final –C), and a relatively high frequency of CCV syllables (Quint 2000a: 35-40, Omnès 1988: 146). However, both insular and continental UGCs also share some features, one of which – the rarity of voiced fricatives – is clearly linked with an African common substrate (Quint 2000a: 112-114).

	Upper Guinea Creoles		
Feature	**Continental**	**Insular**	**Balance**
Prosody (dominant pattern)	all high tone	stress	≠
Dominant syllable-type	CVC	CV	≠
Preferred syllabic structure	monosyllabic	disyllabic	≠
Final –C	pervasive	rare	≠
CCV	-	+	≠
Vowel harmony	+	+	=
Voiced fricatives	rare	rare	=
Lexical reduplication	< 10%	> 25%	≠

Table 9: Comparison of the phonology of insular and continental UGC ideophones.

6.2. Morphosyntax

6.2.1 Word order

Regarding word order, insular UGC ideophones behave exactly like their continental counterparts and immediately follow the element they modify, e.g. (38, 42), except when an object is inserted between a transitive verb and its associated ideophone (45, 46).

(45) Santiago

e	*pegâ-m*	***txápu***	*na*	*brásu*
S3SG	seize.PFV-O1SG	**IDEO**	in/at	arm

'S/he seized **suddenly** my arm **with a firm grip** (lit. s/he seized me **firmly and suddenly** in [the] arm).'

(46) Santiago

e	*abri*	*pórta*	***uandan***
S3SG	open.PFV	door	**IDEO**

'S/he opened the door **wide.**'

6.2.2 Modified elements

As in continental UGCs, insular UGC ideophones usually associate with verbs. However, due to the fact that the distinction between verbs and adjectives is more clear-cut in Capeverdean than in continental UGCs,[36] seven Capeverdean ideophones (out of a total of 34, i.e. 21%) specifically associate with adjectives (as opposed to verbs).[37] Note that only *five* insular (Capeverdean) adjectives, namely *bedju* 'old', *bránku* 'white', *gordu* 'fat', *prétu* 'black' and *ségu* 'blind', have been found to associate with these *seven* ideophones. This difference in number is due to the fact that one and the same adjective can associate with two ideophones. To the best of our knowledge, there is no significant semantic difference between the two ideophones in (47).

(47) Santiago *bedju* 'old'
> *bedju* ***góbu-góbu*** '**very** old'
> *bedju* ***kóti-kóti*** '**very** old'

Furthermore, in Capeverdean, there is at least one case of a fossilized ideophone.

(48) Santiago *sta* 'be' ~ *fika* 'remain' + *nunpriti* 'naked' > *sta/fika nunpriti* 'be/remain naked'

Synchronically, *nunpriti* can be analysed as an adjective. However, it is quite probable that *nunpriti* can be further broken down into *nun* [nũ] 'naked' (a cognate of CC *nuŋ* (see (34)) < Port. *nu* 'naked') + ***priti*** [ˈpriti] (IDEO), a frozen ideophone intensifying the notion of nakedness. In other words, in continental UGCs, the reflex of the Portuguese adjective *nu* 'naked' was reanalysed as an ideophone, whereas in insular

[36] Continental *braŋku* '(be) white', for example, is more appropriately described as a qualifier (or qualifying verb) which can combine with verbal morphemes and be used in the same slot as a verb, whereas Santiago (insular) *bránku* 'white' is not compatible with most verbal morphemes and better described as an adjective in the Portuguese (or Romance) meaning of the term. For more details about the question of adjectives and qualifiers in UGCs, see Biagui et al. (forthcoming), Quint (2000a: 297-300; 2000b: 103; 2023: 457, 459), Biagui (2018: 223-232), Biagui & Quint (2013: 46), Kihm (1994: 34-37, 148-149), Scantamburlo (1981: 44; 1999: 168-169) and Doneux & Rougé (1988: 27, 31, 39-40, 47, 49).

[37] We have classified Santiago Capeverdean *sakédu* 'stand' as a "qualifier" (§5.2.2), as, contrary to most Capeverdean adjectives but similar to its continental counterpart *sikidu* 'stand', it can behave as a verb when used in predicative function (Quint 2000a: 297-300).

UGCs, the reflex of Portuguese *nu* ‘naked’ kept its adjectival status but was intensified by a non-Romance ideophone, *priti*. We may therefore consider that, in Capeverdean, *nun* (in ***nun****priti*) is a sixth instance of an adjective associating with an ideophone.

6.2.3 Morphology

In insular UGC ideophones, expressive reduplication is nonexistent while vowel lengthening and ideophone-to-verb derivation are marginally attested. The lengthening of stressed vowels is documented for color-associated ideophones and has both an attenuative/diminutive and laudatory value (49)-(50).

(49) Santiago *bránku* ‘white’ + *álbu* (IDEO)
> *bránku* ***á****lbu* [ˈalbu] ‘**very** white, **as** white **as snow**’
> *bránku* ***áá****lbu* [ˈaːlbu] ‘**so small, so nice and so** white’

(50) Santiago *prétu* ‘black’ + *finu* (IDEO)
> *prétu* ***fi****nu* [ˈfinu] ‘**very** black, **pitch**-black’
> *prétu* ***fii****nu* [ˈfiːnu] ‘**so small, so nice and so** black’

Some three cases of ideophone-to-verb derivation have been found (51)-(53).

(51) Santiago *sai* ‘get out’ + *futi* (IDEO)
> *sai* ***futi*** [ˈfuti] ‘get out **at once/quickly**’
> *futi* [ˈfuti] ‘get swiftly out [of a place], escape [from a place]’ (e.g. a goat from its pen or a bird from a trap…)

(52) Santiago *labánta* ‘stand up’ + *fakati* (IDEO)
> *labánta* ***fakati*** [fɐˈkɐti] ‘stand up **at once**’
> *fakati****a*** [fɐˈkɐtjɐ] ‘wriggle, move in all directions’

(53) Santiago *gordu* ‘fat’ + *plátxi-plátxi* (IDEO)
> *gordu* ***plátxi-plátxi*** [ˈplatʃi ˈplatʃi] ‘**very** fat, **as** fat **as a pig**’
> *platxi****a*** [ˈplɐtʃjɐ] ‘crash [to the ground]’

As all Santiago insular ideophones were checked for possible derived forms with a trained native-speaker, our list of deideophonic verbs is probably close to exhaustive. Contrary to CC, the derivational process is not morphologically regular: in (51), we have a conversion process, while in (52)-(53), a suffix [ɐ] is added to the ideophone. Note also that, in (53), the reduplication of the ideophonic base is lost in the derived verb. Furthermore, the meaning of the verb is not always predictable. In (52), it is the notion of SUDDEN MOVEMENT that seems to constitute the link between the ideophone and the derived verb, while the common semantic feature in (53) seems to be the SOFT, SHAPELESS CONSISTENCY of the fat (for the ideophone) and of something crashing to the ground and hence losing its original SHAPE and being somewhat SOFTened (for the verb). These cases of ideophone-to-verb derivation are therefore highly idiosyncratic and contrast with the morphologically regular, semantically transparent ideophone-to-verb derivation that occurs in Casamance Creole.

6.2.4 Comparison of the morphosyntactic properties of insular and continental UGCs

As shown in Table 10, the main morphosyntactic differences existing between continental and insular UGCs bear on their respective morphological properties. Generally speaking, continental UGCs display more diverse morphological patterns for their ideophones than insular UGCs.

	Upper Guinea Creoles		
Feature	**Continental**	**Insular**	**Balance**
Word order	V-(O)-IDEO	V-(O)-IDEO	=
Associated elements	mostly verbs	mostly verbs (+adjectives)	=
Expressive reduplication	+	–	≠
Vowel lengthening	intensive	attenuative	≠
Ideophone-to-verb derivation	regular	exceptional	≠

Table 10: Comparison of the morphosyntactic properties of insular and continental UGC ideophones.

6.3. Diatopic variation and inventory

As mentioned in §4.2, only two insular UGC varieties – Fogo and Santiago – have been investigated so far. The main difference between them concerns the size of their inventory: we have found only three ideophones in Fogo vs. 34 in Santiago, and all

Fogo ideophones have close equivalents in Santiago (Table 11). This result comes as no surprise as the African element has already been recognized as clearly less important in Fogo than in Santiago (Quint & Moreira Tavares 2019: 129-131). However, it must also be said that, as the vocabulary of Fogo has not been as thoroughly investigated as that of Santiago, we may have missed some Fogo ideophones.

Fogo			**Santiago cognates**
Modified element	**Ideophone**	**Approximative meaning**	
sta 'be' ~ *fikâ* 'remain'	*enden* [ẽˈdẽ]	'fully empty'	*iondon* [jõˈdõ] ~ *iandan* [jẽˈdẽ] ~ *uandan* [wẽˈdẽ]
kumé 'eat' ~ *bazâ* 'spill' ~ *panhâ* 'catch/seize' ~ *kunprâ* 'buy'	*fépu* [ˈfɛpu]	'completely, without leaving anything'	*fépu* [ˈfɛpu]
prétu 'black'	*finu* [ˈfinu]	'very'	*finu* [ˈfinu]

Table 11: The three known Fogo ideophones with their Santiago counterparts.[38]

It is also possible that more ideophonic items will be found on the two other islands of Sotavento (or Southern Capeverdean, see §2.1 and Map 2), namely Brava and Maio, which have not been explored for this matter. Furthermore, we expect that ideophones, which seem to be mostly a Niger-Congo feature in UGCs (see §5.4 and §6.4), are much less common in Barlavento (or Northern Capeverdean), where lexical African influence is weaker than in Sotavento (see §4.2).

6.4. Origin and comparison with continental UGCs

6.4.1 Origin of the insular UGC ideophones

Seventeen insular UGC ideophones (i.e. 17/34 = 50%) can be traced back to Mandinka and/or Wolof (Table 12). Basically, the ratio of Mandinka and Wolof is similar to what has been observed in continental UGCs (see §5.4 and Table 6), and most of the African-derived insular UGC ideophones have cognates in continental

[38] For Fogo, *enden* is attested in Moreira Tavares (2020: 237), *fépu* in Moreira Tavares (2020: 273) and Quint & Moreira Tavares (2019: 152), and *finu* in Quint & Moreira Tavares (2019: 148).

UGCs. If we consider the distribution types (with due caution because of the limited amount of data), Mandinka seems to predominate in the 'common core' type while undoubtedly Wolof-derived ideophones appear only in the 'CC-only' and 'minority models' types. Here again, the distribution profile is very similar to what has been observed for continental UGC African-derived ideophones (ADI).

	Origin				
Distribution type[39]	**Wolof**	**Mandinka**	**Wolof or Mandinka**	**Total ADI**	**Continental cognates**
CC-only	**1**	0	0	1	1
Minority models	2	**3**	1	6	4
Common core	0	**3**	3	6	5
Other distribution	0	1	**2**	3	2
NA[40]	**1**	0	0	1	1
Total	**4**	**7**	**6**	**17**	**13**

Table 12: African-derived insular UGC ideophones according to their distribution type as defined in Table 4.

Like on the continent, some insular UGC ideophones come from Portuguese (54) or can be derived both from an African language and/or Portuguese (55).

(54) Santiago *bránku* 'white' + *álbu* (IDEO) > *bránku* ***álbu*** '**very** white, **as** white **as snow**', with *álbu* < Portuguese *alvo* 'bright/intense white'

(55) Santiago *prétu* 'black' + *finu* (IDEO) > *prétu* ***finu*** '**very** black, **pitch**-black', with *finu* < Mandinka *fíŋ* 'be black' and/or Portuguese *fino* 'fine, refined'

[39] The distribution types are those defined for continental UGCs, i.e. if a given insular UGC ideophone X is the cognate or semantic equivalent of a continental ideophone Y, X is ascribed to the distribution type Y belongs to.

[40] Some insular ideophones have no attested (NA) cognate or semantic equivalent in the list of 81 Casamance Creole (CC) ideophones which has been used as the standard for continental UGCs (§4.1). However, in some cases, we were able to find afterwards a continental cognate or semantic equivalent outside the CC list, e.g. Santiago *pila* 'crush' ~ *kebra* 'break' + *moku* (IDEO) > *pila/kebra* ***moku*** 'break/crush **completely**'. After *moku* was found in Santiago Capeverdean, it was also found with a similar meaning and form both in Casamance and Bissau3 continental varieties. Furthermore, *moku* can be traced to a Wolof term, *mokk* [mɔːkkə] 'be ground/crushed' (Diouf 2003: 226; Dieng 1985: 247). But as *moku* was not included in the initial list of 81 CC ideophones, it is counted in Table 11 as an insular African-derived ideophone belonging to the 'non-attested (NA)' type.

6.4.2 Comparison with continental UGCs

The first difference between insular and continental UGC ideophones concerns their number: there are less ideophones on the islands than on the continent. Only 31 ideophones[41] from insular varieties correspond to the 81 ideophones of the standard Casamance Creole list. Put differently, 50 ideophones of the standard CC list lack an equivalent in Santiago Capeverdean, the only insular variety to have been extensively checked. The number of missing equivalents is thus much higher than for all continental varieties that have been checked with consultants, where there are between 23 and 36 gaps per list (see §5.3.2 and Table 5).

Admittedly, the standard CC list is probably far from exhaustive, and there must be many more ideophones, both in continental and insular varieties, that were not taken into account in the present study. However, from our own experience of both UGC branches (including daily practice thereof), we can say quite confidently that the ideophone category is less developed in insular than in continental varieties. As the category seems to draw most of its members from the African substrate and adstrate (see 5.4.), which is stronger on the continent than on the archipelago of Cape Verde, the difference in the number of ideophones seems to be consistent with what we know about the history of UGCs.

When we take the distribution types defined for continental UGC ideophones (see §5.3.1 and Table 4) and apply them to three subsets of our insular UGC sample of ideophones (Table 13), namely (i) the semantic equivalents of CC ideophones, (ii) the cognates of CC ideophones and (iii) the African-derived ideophones (ADI), the most striking observation is the similarity between continental and insular varieties regarding the ratio of each distribution type, no matter which subset is taken into account. For all subsets, the items belonging to the continental 'common core' type make up approximately one half of the total number of insular ideophones, just as in CC. All in all, insular and continental UGC ideophones share the same distribution profile.

[41] We have excluded the three insular ideophones belonging to the 'NA' type (see fn. 40), as they are precisely defined by the fact that they have no CC equivalent. Hence the value of 31 insular UGC ideophones (= 34 (total) – 3 (NA)) which we consider for the comparison with continental UGCs.

		Insular (Santiago) ideophones		
Distribution type	**CC**	**Semantic equivalents**	**Continental cognates**	**ADI**
CC-only	22	4	3	1
Minority models	14	7	5	6
Common core	**39**	**15**	**9**	**6**
Other distribution	9	5	3	3
NA	0	3	1	1
Total	84	34	21	17

Table 13: Comparison of continental (CC) UGC ideophones with different subsets of insular UGC ideophones according to their distribution type as defined in Table 4.

However, if we consider the semantic types of insular UGC ideophones (Table 14), a difference appears between the two UGC branches. Regarding the total sample of insular UGC ideophones, the semantic distribution is quite comparable to what has been observed for continental varieties, with a clear predominance of intensifying ideophones: 25 out of 34, i.e. 74%, a value equivalent to what was found in the CC-standard list (75%, see §4.3). Nonetheless, if we take into account, for each semantic type, the number of insular UGC ideophones that have a plausible cognate in continental UGCs, a clear split appears:

- Almost all (8/9) insular UGC non-intensifying ideophones have cognates in continental UGCs.
- Only one half (13/25) of insular UGC intensifying ideophones have cognates in continental UGCs.

In other terms, insular intensifying ideophones seem to be less easily traceable to proto-UGC than ideophones of other semantic categories. Yet it appears that it is precisely this semantic category that was favoured and considered as the most prototypical during the elaboration of the standard CC list (§4.3). Why then are intensifying ideophones less prone to be shared between continental and insular UGCs? The main underlying reason may well be once more the respective degree of Niger-Congo influence. As a matter of fact, intensifying ideophones are certainly a salient semantic category when contrasting Niger-Congo languages with Portuguese, the main lexifier of both insular and continental UGCs (§2.2). Whereas, for instance, sound ideophones (or “onomatopoeias”) are found in Portuguese (and in other Romance languages), a word category matching Niger-Congo intensifying ideophones

is missing. This may well explain the relative lack of stability of intensifying ideophonic roots in insular UGCs: these varieties, submitted to a stronger Portuguese influence, must have aligned their semantic organization with their Romance lexifier and lost African features (such as the use of intensifying ideophones) that contrasted too drastically with Portuguese. In contrast, continental UGCs, which had less intense contacts with Portuguese (and more African influence), were more prone to retain Niger-Congo material.

Semantic type		**Total**	**Continental cognates**
Intensifying		**25**	**13**
Non-intensifying		**9**	**8**
of which:	Sound (onomatopoeia)	4	4
	Immediate	3	3
	Others	2	1
Total		**34**	**21**

Table 14: Semantic types of insular UGC ideophones.

The semantic field of color provides a good example of the difference between insular and continental ideophones regarding intensifying ideophones. Many Niger-Congo languages make a lexical distinction between three (more rarely four, e.g. Koalib) basic colors ('white' = FULL LIGHT, 'red' = FULL COLOR and 'black' = NO LIGHT, NO COLOR, see Segerer & Vanhove (2019: 292, 315-316); Rougé (1988: 40-41)), each of which is intensified by a specific ideophone (Table 15). These color intensifiers are among the most stable ideophones, and it is possible to reconstruct proto-forms for some branches of Niger-Congo (for Kordofanian, see examples in Quint 2018: 202-204).

Intensifying ideophone						
	Niger-Congo languages				**UGCs**	
Color	**Mandinka**	**Wolof**	**Djifanghor Nyun**	**Koalib**	**Proto-conti-nental UGC**	**Insular**
'white'	*fér* [+*léw* ~ *wéj*]	*fur* ~ *tàll*	*pat*	*ppɛ́ppɛ̀r*	*/nok/	[ˈalbu]
'red'	*cáw* ~ *táw* ~ *táráw*	*coy* ~ *cur(r)*	*úutut*	*ccél-ccél* ~ *ccéccél*	*/wak/	—
'black'	*kím* ~ *kírím* [+*hót* ~ *nót*]	*kuk(k)*	*ñɔtɔtɔ́t*	*ʈíʈʈít*	*/fandaŋ/	[ˈfinu]

Table 15: Intensifying ideophones for the three basic coulours 'white, red, black' in several Niger-Congo languages, in proto-continental and in insular UGCs.

Table 15 shows that the system of three color-intensifying ideophones of continental UGCs fits in with the general Niger-Congo semantic pattern. In contrast, insular UGCs differ from the Niger-Congo (and continental UGC) model in two respects: They lack an intensifier for 'red', and they have developed Romance (or partially Romance)-derived ideophones — see (54) and (55) – to intensify 'black' and 'white', whereas continental UGCs have African (or at least non-Romance) intensifying ideophones for the same colors.

In sum, the semantic field of color illustrates quite well the difference between the system of continental UGC ideophones, which has stronger semantic and phonological links with Niger-Congo languages, and the system of insular UGC ideophones, whose African-derived component, albeit still visible, is weaker, both phonologically and semantically, than in continental UGCs.

6.4.3 Comparison of the origin and semantic behavior of insular and continental UGC ideophones

Table 16 summarizes the main commonalities and differences between continental and insular UGC ideophones regarding their origin and semantic organization.

	Upper Guinea Creoles		
Feature	**Continental**	**Insular**	**Balance**
Origin	Mandinka + Wolof	Mandinka + Wolof	=
Inventory	> 40	< 40	≠
Dominant distribution type	common core	common core	=
Intensifying ideophones: cognates with continental UGCs	100%	50%	≠

Table 16: Comparison of the origin and semantic organization of insular (=Capeverdean) and continental UGC ideophones.

7. Conclusion

This paper has described in great detail the phonology, morphosyntax, diatopic variation, origin and semantic properties of UGC ideophones. It has also provided a systematic comparison of the characteristics of this particular word class in two branches of UGCs: (i) continental (Casamance and Guinea-Bissau) and (ii) insular

(Capeverdean) UGCs. In our study, we were able to identify and define several tens of ideophones in continental and insular UGCs. The existence of these ideophones seems to be mainly due to the influence exerted by the Niger-Congo African substrate (for both insular and continental UGCs) or/and adstrate (continental UGCs only). The observation that many ideophonic roots are shared by continental and insular UGCs reinforces the hypothesis of a common stage (the proto-Upper-Guinea-Creole or proto-UGC) from which both UGC branches are descended. The fact that various African-derived ideophones can be traced back to proto-UGC implies that this common stage must have been different from Portuguese (§2.2), a Romance language which does not have an ideophone category. Yet, despite obvious resemblances, continental and insular UGC ideophones also reflect a clear split between the two branches, as the influence of Niger-Congo languages is significantly heavier on the continent. This important result is confirmed by other similar findings (Rougé 1999, Quint 2000b: 99-117, Quint & Moreira Tavares 2019: 126-127) regarding the respective influence of the African substrate on insular and continental UGCs.

In the future, the present research could be expanded in several directions. Increasing the sample of ideophones would improve our knowledge of all continental and insular varieties. In particular, the three Sotavento (Southern Capeverdean) varieties other than Santiago should be thoroughly investigated, as we have almost no data for Brava and Maio and too few for Fogo. Furthermore, our assessment of the substrate and/or adstrate influence would be strengthened if more Niger-Congo languages were checked. Balant (bjt/ble), Jola, Manjak (mfv), Papel (pbo) and Nyun, which are all spoken in direct contact with one or several of today's continental UGCs, probably have a lot to offer in this respect. Finally, a refinement and an elaboration of the semantic and distribution classifications that have been proposed in this paper seems desirable. Large, accessible, searchable corpora for the UGCs would allow us to pay more attention to the collocations involving ideophones and their actual usage in spontaneous speech.

It is clear that the present study is far from having exhausted the question of UGC ideophones, which remain a most exciting topic for anyone interested in language contact effects on ideophones, in general, and the grammar and genesis of the Afro-Portuguese varieties of Cape Verde, Casamance and Guinea-Bissau, in particular.

Acknowledgements

We would like to thank Aires Semedo and Karina Moreira Tavares for their comments and advice on Capeverdean data, Braima Sam Mendes for his help during our fieldwork in Guinea-Bissau, and Jean-Baptiste Maistre for his bibliographical support. Our thanks also go to all the native speakers who were kind enough to share with us their knowledge of their respective varieties. We also are grateful to the anonymous reviewers of this paper for their helpful suggestions and to special issue editors for their meticulous editing.

Abbreviations

1 = 1st person	IDEO/IDEO = ideophone	PL = plural
2 = 2nd person	IPFV = imperfective	RED = reduplicated
3 = 3rd person	NA = non-attested	S = stressed
ADI = African-derived ideophone	NEG = negation	S/S = subject
ASS = assertive	O/O = object	SG = singular
CC = Casamance Creole	PFV = perfective	SUB = subordinator
DEM = demonstrative	Port. = Portuguese	U = unstressed
F = feminine	preC = preceding consonant	UGC = Upper Guinea Creole
H = high tone	PROG = progressive	
HAB = habitual	PROX = proximal	

References

Biagui, Noël B., Jean-François Nunez & Nicolas Quint. (forthcoming). *Casamance Creole*. In Lüpke, Friederike (ed.), *Oxford guide to the world's languages: Atlantic*. Oxford: Oxford University Press.

Biagui, Noël B. & Nicolas Quint. 2013. *Casamancese Creole*. In Susanne Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Survey of pidgin and creole languages*, Vol. 2: *French-based, Portuguese-based and Spanish-based languages*, 40–49. Oxford: Oxford University Press.

Biagui, Noël B. 2018. *Description générale du créole afro-portugais de Ziguinchor(Sénégal)*. Paris: Karthala.

Biagui, Noël B. (in preparation). *Dictionnaire créole casamançais-français*.

Childs, Tucker. 1994. Expressiveness in contact situations: The fate of African Ideophones. *Journal of Pidgin and Creole Languages* 9(2). 257–282.

Couto, Hildo H. 1994. *O Crioulo Português da Guiné-Bissau*. Hamburg: Helmut Buske.

Couto, Hildo H. 1995. Exclusive particles (ideophones) in Guinea-Bissau creole. In Philip Baker (ed.), *From contact to creole and beyond*, 207–215. London: University of Westminster Press.

Creissels, Denis. 1994. *Aperçu sur les structures phonologiques des langues négro-africaines*. Grenoble: ELLUG.

Creissels, Denis. 2006. *Syntaxe structurale, une introduction typologique 1. Catégories et Construction*. Paris: Lavoisier.

Creissels, Denis. 2012. *Lexique mandinka-français*. Unpublished MS. Available online at: http://www.deniscreissels.fr/public/Creissels-lexique_mandinka_2012.pdf (Accessed 28/03/2023).

Creissels, Denis. 2013. *Le mandinka – phonologie, grammaire, textes*. Paris: Karthala.

Creissels, Denis, Sidia Jatta & Jobarteh Kalifa. 1982. Lexique mandinka-français. *Mandenkan* 3.

Dalphinis, Morgan. 1986. *African language influences in creoles lexically based on Portuguese, English and French, with special reference to Casamance Kriul* [sic]*, Gambian Krio and St. Lucian Patwa*, Vol. 1. Doctoral dissertation, University of London.

Dieng, Sady. 1985. *Étude contrastive des lexies du créole endogène "portugais" et du wolof*. Doctoral dissertation, Université Lumière-Lyon 2.

Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6/10. 654–672.

Dingemanse, Mark. 2019. 'Ideophone' as a comparative concept. In Kimi Akita & Prashant Pardeshi (eds.), *Ideophones, mimetics and expressives*, 13–34. Amsterdam: John Benjamins.

Diouf, Jean-Léopold. 2003. *Dictionnaire wolof-français et français-wolof*. Paris: Karthala.

Doneux, Jean-Léonce & Jean-Louis Rougé. 1988. *En apprenant le créole à Bissau ou Ziguinchor*. Paris: L'Harmattan.

Fal, Arame, Rosine Santos & Jean-Léonce Doneux. 1990. *Dictionnaire wolof-français, suivi d'un index français-wolof*. Paris: Karthala.

Jacobs, Bart. 2012. *Origins of a creole. The history of Papiamentu and its African ties*. Berlin: De Gruyter.

Kihm, Alain. 1994. *Kriyol syntax: The Portuguese-based creole language of Guinea-Bissau*. Amsterdam: John Benjamins.

Mbodj, Chérif. 1979. *Phonologie du créole de Guinée-Bissau*. Dakar: Centre de linguistique appliquée de Dakar.

Mbodj, Chérif. 1984. *Grammaire du créole de Guinée-Bissau*. Bissau: Unpublished MS.

Moreira Tavares & Ana Karina. 2020. *Documentation et description grammaticale et lexicale du créole afro-portugais de l'île de Fogo (République du Cap-Vert, Afrique de l'Ouest) / Documentação e descrição gramatical e lexical do crioulo afro-português da ilha do Fogo (República de Cabo Verde, África Ocidental)*. Doctoral dissertation, INALCO, Paris.

Nuckolls, Janis B., Joseph A. Stanley, Elizabeth Nielsen & Roseanna Hopper. 2016. The systematic stretching and contracting of ideophonic phonology in Pastaza Quichua. *International Journal of American Linguistics* 82(1). 95–116.

Omnès, Robert. 1988. Les structures syllabiques du castillan, étude synchronique et diachronique, comparaison avec 28 autres langues. In *Actes du 3e colloque de linguistique hispanique*, 135–155. Grenoble: Université Stendhal-Grenoble 3.

Quint, Nicolas. 1998. *Dicionário de caboverdiano-português*. Lisbon: Verbalis.

Quint, Nicolas. 1999. *Dictionnaire cap-verdien-français*. Paris: L'Harmattan.

Quint, Nicolas. 2000a. *Grammaire de la langue capverdienne (description diachronique et synchronique du créole afro-portugais des îles du Cap-Vert)*. Paris: L'Harmattan.

Quint, Nicolas. 2000b. *Le cap-verdien: origines et devenir d'une langue métisse*. Paris: L'Harmattan.

Quint, Nicolas. 2006. *Phonologie de la langue koalibe*. Paris: L'Harmattan.

Quint, Nicolas. 2008. *L'élément africain dans la langue capverdienne / Africanismos na língua caboverdiana*. Paris: L'Harmattan.

Quint, Nicolas. 2009a. *The phonology of Koalib, a Kordofanian language from the Nuba Mountains (Sudan)*. Cologne: Rüdiger Köppe.

Quint, Nicolas. 2009b. O cabo-verdiano: uma língua mundial. *Revista de Estudos Cabo-Verdianos* 3. 129–144.

Quint, Nicolas. 2018. Koalib ideophones: a first overview of their main characteristics. In Gertrud Schneider-Blum, Birgit Hellwig & Gerrit J. Dimmendaal (eds.), *Nuba Mountain languages studies: new insights*, 175–207. Cologne: Rüdiger Köppe.

Quint, Nicolas. 2023. Guinea-Bissau. In Ursula Reutner (ed.), *Manual of Romance languages in Africa*, 447–465. Berlin: De Gruyter.

Quint, Nicolas & Karina Moreira Tavares. 2019. The common African lexical core of Upper Guinea Creoles and its historical significance. *Journal of Ibero-Romance Creoles* 9. 115–161.

Rougé, Jean-Louis. 1988. *Petit dictionnaire étymologique du kriol de Guinée-Bissau et de Casamance*. Bissau: Instituto Nacional de Estudos e Pesquisa.

Rougé, Jean-Louis. 1999. Apontamentos sobre o léxico de origem africana dos crioulos da Guiné e de Cabo Verde (Santiago). In Klaus Zimmermann (ed.), *Lenguas criollas de base lexical espanola y portuguesa*, 49–67. Frankfurt am Main: Vervuert.

Rougé, Jean-Louis. 2004. *Dictionnaire étymologique des créoles portugais d'Afrique*. Paris: Karthala.

Scantamburlo, Luigi. 1981. *Gramática e dicionário da língua criol da Guiné-Bissau (GCr)*. Bologna: Editrice Missionaria Italiana.

Scantamburlo, Luigi. 1999. *Dicionário guineense,* Vol. 1: *Introdução e notas gramaticais*. Lisbon: Edições Colibri/Faspebi (Fundação para o Apoio ao Desenvolvimento dos Povos do Arquipélago de Bijagós).

Scantamburlo, Luigi. 2002. *Dicionário do guineense*, Vol. 2: *Dicionário guineense-português*. *Disionariu guinensi-purtuguis*. Lisbon: Colibri/Faspebi.

Segerer, Guillaume & Martine Vanhove. 2019. Color naming in Africa. In Ida Raffaelli, Daniela Katunar & Barbara Kerovec (eds.), *Lexicalization patterns in colour naming: a cross-linguistic perspective*, 287–330. Amsterdam: John Benjamins.

Truppi, Chiara & Patrícia Costa. 2019. *On ideophones in Portuguese Creoles*, paper presented at *Societas Linguistica Europaea – 52nd Annual Meeting*, 21st–24th August 2019 in Leipzig.

Wilson, William A. A. 1962. *The Crioulo of Guiné*. Johannesburg: Witwatersrand University Press.

CONTACT

nicolas.quint@cnrs.fr
biaguinoelbernard@yahoo.fr

Appendices

Appendix 1: Continental UGC ideophones classified according to their distribution type

Table 1.1: 'CC-only' type.

	CC ideophone					Associated element			Other continental UGC equivalents
No.	**Ideophone**	**SS**	**Etymon**	**SEM**	**RED**	**Form**[42]	**English**	**PoS**	
(1)	*bok-bok*	CVC-CVC	—	intens.	Yes	*termé*	'shiver'	v.	—
(2)	*ca*	CV	—	intens.	No	*fritá*	'fry'	v.	—
(3)	*cañar*	CVCVC	—	intens.	No	*jugtá ~ jugutá*	'jump'	v.	*caŋ* (+ *saltá* 'jump') (Cacheu3)
(4)	*cip*	CVC	—	intens.	No	*disí*	'get down'	v.	*cɐm ~ caŋ* (Cacheu2)
(5)	*colop*	CVCVC	Mandinka©	intens.	No	*jugtá ~ jugutá*	'jump'	v.	*cas* (Cacheu1 + Cacheu3); *cɒm ~ caŋ* (Cacheu2)
(6)	*fes*	CVC	—	intens.	No	*simí*	'disappear'	v.	—
(7)	*fututut*	CVCVCVC	—	intens.	Yes	*kerá*	'smell'	v.	—
(8)	*kɐc*	CVC	—	intens.	No	*jundá*	'drink'[43]	v.	*forgot* (Cacheu3)
(9)	*kɐkɐt*	CVCVC	—	intens.	Yes	*risu*	'be hard/tough'	qual.	—

[42] As Casamance Creole is the standard for the comparison of continental UGCs (§2.1), the form of the associated element is the one attested in CC, unless otherwise specified.

[43] The basic meaning of *jundá* is 'pull'. However, *kis* intensifies a figurative meaning of *jundá*, namely 'drink (alcohol)'.

	CC ideophone					Associated element			Other continental UGC equivalents
No.	**Ideophone**	**SS**	**Etymon**	**SEM**	**RED**	**Form**[42]	**English**	**PoS**	
(10)	*kamaɟ*	CVCVC	Wolof©	immdt.	No	*pagá*	'switch off'	v.	*fup (?)* (Cacheu3); *vut* (Bissau1)
(11)	*kik*	CVC	—	intens.	No	*sikidu*	'be straight' (pole/stake)	qual.	—
(12)	*fof*	CVC	Mandinka©	intens.	No	*lebi*	'be light'	qual.	*kef* (Cacheu3)
(13)	*koŋ*	CVC	Wolof©	intens.	No	*risu, seku*	'be hard/tough, be dry'	qual.	—
(14)	*mɐp*	CVC	—	intens.	No	*barí*	'get, win, collect'	v.	*fep* (Cacheu2); *bup* (Cacheu3)
(15)	*mes*	CVC	Wolof©	intens.	No	*perdé*	'disappear'	v.	*buk* (Cacheu3)
(16)	*pɐɟɐɟ*	CVCVC	—	speech ('clearly')	No	*kontá*	'tell'	v.	—
(17)	*paraw*	CVCVC	—	intens.	No	*camí*	'get drunk'	v.	*tɔrɔk* (Cacheu2)
(18)	*pip*	CVC	—	speech ('threateningly')	No	*falá*	'say'	v.	—
(19)	*pul*	CVC	Wolof©	immdt.	No	*sey*	'get out'	v.	*fas* (Cacheu2); *fat* (Bissau1)
(20)	*rat*	CVC	—	intens.	No	*dá*	'drink'	v.	—
(21)	*tɐkɐp*	CVCVC	—	intens.	No	*pañá*	'thicken'	v.	—
(22)	*was*	CVC	Mandinka + + Wolof©	intens.	No	*pajigá*	'scatter'	v.	*buk* (Cacheu3); *fep* (Bissau1)

Table 1.2: 'minority model' type.

Continental UGC common ideophone								Associated element			Other continental UGC equivalents
No.	**Distr. profile**	**Ideophone**	**Attested forms**	**SS**	**Etymon**	**SEM**	**RED**	**Form**	**Meaning**	**PoS**	
(1)	CC+2	*/bap/	*bap* (CC); *vap* (Bissau1); *wap* (Bissau3)	CVC	Wolof©	intens. +immdt.	No	*sintá/kay*	'sit, fall'	v.	*rip* (Cacheu3)
(2)	CC+2	*/fas/	*fas* (CC+Cacheu1); *fas (?)* ~ *fes* (+ *rapá* 'shave') ~ *feles* (Geba)	CVC	Wolof©	intens.	No	*limpu*	'be clean'	qual.	*wañja(ŋ)* ~ *wac (?)* (Geba)
(3)	CC+3	*/fuf/ ~ */fus/	*fuf* (CC+Bissau3); *fus* (Cacheu2+Bissau1)	CVC	Wolof©	intens.	No	*mas*	'be more'	v.	—
(4)	CC+1	*/kef/	*kef* (CC+Cacheu1)	CVC	—	intens.	No	*fartá*	'be satisfied' (eating)	v.	*diŋ* ~ *dir* (Geba); *kuŋ* (Geba+Cacheu2);[44] *lip* (Cacheu3)
(5)	CC+3	*/kokorot/	*kokorot* (CC+Cacheu1+ Cacheu3+Bissau1)	CVCVCVC	—	intens.	Yes	*beju*	'be old'	qual.	—
(6)	CC+2	*/las/	*las* (CC+Bissau1); *lac* (Cacheu1)	CVC	Wolof©	intens.	No	*ditá*	'lie'	v.	*caŋ* (+ lastrá 'stretch out') (Geba)
(7)	CC+1	*/net/	*net* (CC+CA2)	CVC	—	intens.	No	*pí*	'put'	v.	—
(8)	CC+1	*/pirkit/	*pirkit* (CC+Geba)	CVCCVC	Mandinka©	immdt.	No	*labantá*	'stand up'	v.	—
(9)	CC+3	*/puf/	*puf* (CC+Geba+Cacheu3+	CVC	—	intens.	No	*furá*	'make a hole,	v.	*fut* (Cacheu3)

[44] As Geba and Cacheu2 have the same form (*kuŋ*), we could have considered this example as an illustration of the CC vs. Guinea-Bissau subtype of the 'other distribution' type (§3.3.1 and Table 1.4). However, the ideophone */kuŋ/ is also attested in at least seven different continental UGCs as intensifying *yiñcí* 'be full' (Table 1.3). Hence, it seems that the form *kuŋ* attested here in Geba and Cacheu2 is due to a confusion between being 'full' and 'satisfied' (= 'full of food') rather than to a genuine semantic peculiarity shared by Geba and Cacheu2.

Continental UGC common ideophone								Associated element			Other continental UGC equivalents
No.	**Distr. profile**	**Ideophone**	**Attested forms**	**SS**	**Etymon**	**SEM**	**RED**	**Form**	**Meaning**	**PoS**	
			Bissau1)						pierce’		
(10)	CC+3	*/tar/	*tar* (CC+Geba+ Cacheu3+ Bissau1); *tax* [tax] (Geba); *par* (Cacheu1)	CVC	—	intens. +onom.	No	*bafatiyá*	‘slap’	v.	*baw* (Geba); *paw* (Cacheu1+Cacheu2)[45]; *fak* (Cacheu1); *fap* (Bissau3); *vap* (Cacheu3)
(11)	CC+3	*/tep/	*tep* (CC+ +Bissau3); *tep* (+ *fartá* ‘be satisfied) (Cacheu1); tip ((+ *fartá*) (Cacheu3)	CVC	Mandinka©	intens.	No	*yiñcí*	‘fill’ ~ ‘be full’	v.	—
(12)	CC+1	*/tim/	*tim* (CC+Geba)	CVC	—	onom. (‘hard surface’)	No	*kay*	‘fall’	v.	*fat* (Cacheu1); *kaŋkalaŋ* ~ *puf* (Cacheu2); *tuc* (Cacheu3); *tup* (Bissau1)
(13)	CC+2	*/wac/	*wac* (CC); *wap* (Cacheu1); *vac* (+ *yabrí* ‘open’) (Bissau1)	CVC	Mandinka©	intens.	No	*fendé*	‘split’	v.	*parac* ~ *tip* (+ *yabrí* ‘open’) (Bissau1)
(14)	CC+2	*/jambalaŋ/ ~ */wandalaŋ/	*yambalaŋ* (CC); *wandalaŋ* (Cacheu1+Cacheu2)	CVCCVCVC	Wolof©	visual (‘desert’)	No	*sá (CC) (=sta)/fiká*	‘be’	v.	—

[45] Here also, as for */kef/ above, we have come to the conclusion that the common form *paw* found in Cacheu1 and Cacheu2 does not necessarily illustrate a semantic peculiarity shared by these two varieties. The sound produced by a slap is probably */tar/ across most of the UGC continental varieties, but some consultants resorted to other, less specialized sound ideophones such as *paw*, which we have found in other varieties in quite different contexts (e.g. to describe the noise produced by a broken stick in CC).

Table 1.3: 'common core' type.

Continental UGC common ideophone								Associated element			Other CUGC equiv.
No.	**Distr. profile**	**Ideophone**	**Attested forms**	**SS**	**Etymon**	**SEM**	**RED**	**Form**	**Meaning**	**PoS**	
(1)	CC + 7	*/baw/[46]	*baw* (CC + Geba + Cacheu1 + Cacheu2 + Bissau2); *badaw* (Cacheu3 + Bissau1 + Bissau3); *bak* (Bissau2)	CVC	—	intens.	No	*forti*	'be sour'	qual.	—
(2)	CC + 4	*/bik/	*bik* (CC + Geba + Cacheu1 + Bissau1)	CVC	—	intens.	No	*kabá*	'be finished'	v.	—
(3)	CC + 6	*/buk/	*buk* (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	—	intens.	No	*kabá, pajigá*	'be finished, scatter'	v.	—
(4)	CC + 4	*/bup/	*bup* (CC + Bissau1); *vup* (Geba); *bip* (Cacheu1); *vip* (Cacheu1 + Cacheu3)	CVC	—	onom.	No	*kay*	'fall'	v.	—
(5)	CC + 7	*/caŋ/ ~ */kaŋ/[47]	*caŋ* (Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau2 + Bissau3); *kaŋ* (CC + Cacheu1 + Cacheu3 + Bissau1); *taŋ* (Bissau3)	CVC	—	intens.	No	*firmá/sikidu*	'stand' (person)	v./ qual.	—
(6)	CC + 4	*/cap/	*cap* (CC + Geba + Cacheu1 + Cacheu3 + Bissau1) ~ *ñap* [ɲap] (Geba)	CVC	Mandinka©	intens.	No	*pañá/pegá*	'catch'	v.	—
(7)	CC + 4	*/corot/	*corot* (CC + Cacheu1); *corot* (+ *pikininu* 'be small') (Geba + Cacheu2) *ñeret*	CVCVC	Mandinka©	atten.	No	*pañá, partí*	'take, give'	v.	—

[46] The form */baw/ has been retained as the base form for this ideophone against *badaw*, which is almost certainly an extended (expressive) form of *baw*.

[47] The form */kaŋ/ is obviously the same as the one that is used to intensify *tesu* 'be solid' (see Table 1.4). Some continental UGC varieties seem to use the same form to intensify 'be solid' and 'stand', while others use one specific ideophone for each of these two meanings.

Continental UGC common ideophone								Associated element			Other CUGC equiv.
No.	**Distr. profile**	**Ideophone**	**Attested forms**	**SS**	**Etymon**	**SEM**	**RED**	**Form**	**Meaning**	**PoS**	
			(+*pikininu*) (Cacheu3)								
(8)	CC+5	*/culup/	*culup* (Geba+Cacheu1+Cacheu3+Bissau1); *culum* (CC+Cacheu2)	CVCVC	—	onom.	No	*fundá*	'dive, sink'	v.	—
(9)	CC+5	*/cut/	*cut* (CC+Cacheu1+Cacheu2+Cacheu3+Bissau1+Bissau3)	CVC	—	intens.	No	*melá*	'be sweet'	v.	—
(10)	CC+5	*/fa(r)kat/ ~ */fɛ(r)kɛt/	*fɛkɛt* (CC); *fakat* (Cacheu1); *fɛrkɛt* ~ *ferket* (Geba); *fargat* (Bissau1); *fat* (Cacheu2); *fat* (+ kordá 'wake up') (Bissau1); *firgit* (Cacheu3+Bissau2)	CVCVC	Mandinka+Wolof©	immdt.	No	*labantá*	'stand up'	v.	*caŋ* (Geba)
(11)	CC+6	*/fandaŋ/	*fandaŋ* (CC+Geba+Cacheu1+Cacheu2+Cacheu3+Bissau1+Bissau3)	CVCCVC	—	intens.	No	*braŋku*	'be white'	qual.	—
(12)	CC+6	*/fep/	*fep* (CC+Cacheu1+Cacheu2+Cacheu3+Bissau1+Bissau3); *fɛp* (Geba)	CVC	Mandinka+Wolof© (QM)	intens.	No	*kabá*	'finish'	v.	—
(13)	CC+7	*/fit/	*fit* (CC+Geba+Cacheu1+Bissau1+Bissau2+Bissau3); *fis* (Cacheu2); *fif* (Cacheu3); *fas* (Bissau2); *falas* (Geba)	CVC	Mandinka©	immdt.	No	*pasá*	'go by'	v.	—
(14)	CC+5	*/fitfat/	*fitfat* (CC+Geba+Cacheu1+Cacheu2+Cacheu3+Bissau1)	CVCCVC	Mandinka©	intens.+onom.	Yes	*jantí*	'be hurried'	v.	—
(15)	CC+6	*/fut/	*fut* (CC+Geba+Cacheu1+Cacheu2+Cacheu3+Bissau3); *fat*	CVC	Mandinka+Wolof©	immdt.	No	*sey*	'get out'	v.	—

Continental UGC common ideophone								Associated element			Other CUGC equiv.
No.	**Distr. profile**	**Ideophone**	**Attested forms**	**SS**	**Etymon**	**SEM**	**RED**	**Form**	**Meaning**	**PoS**	
			(Cacheu3 + Bissau1); *fup* (Bissau3); *fis* (Bissau3)								
(16)	CC + 6	*/jɔp/	*yɔp* (Geba + Cacheu1 + Cacheu2 + Bissau1); *yop* (CC + Cacheu3 + Bissau3)	CVC	Wolof©	intens.	No	*mojá*	'get wet'	v.	*pɔcɔk* (Geba)
(17)	CC + 6	*/jem/	*yem* (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	Mandinka©	intens.	No	*friyá*	'be cold'	qual.	—
(18)	CC + 6	*/kaŋ/	*kaŋ* (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1); *kaŋ* (+ *seku* 'be dry) (Bissau3); *kaŋkaraŋ* (Bissau1)	CVC	Wolof©	intens.	No	*risu*	'be hard/ tough'	qual.	—
(19)	CC + 4	*/kat/	*kak* (CC); *kat* (Cacheu1 + Cacheu3); *pakat/ fakat* (Geba); *pargat* (Cacheu2); *fargat* (Cacheu3)	CVC	—	intens.	No	*kebrá*	'break'	v.	—
(20)	CC + 6	*/keŋkereŋ/	*keŋkereŋ* (CC + Geba + Cacheu1 + Cacheu3 + Bissau1 + Bissau3); *kereŋ* (Cacheu2); *keŋ* (Geba)	CVCCVCVC	—	intens.	Yes	*saŋ*	'be healthy'	qual.	—
(21)	CC + 6	*/kondoŋ/	*kondoŋ* (CC + Geba + Cacheu1 + Cacheu2 + Bissau3); *kindiŋ-kondoŋ* (Cacheu3 + Bissau1); *konkoroŋ* (Geba)	CVCCVC	—	atten. 'completely alone'	No	*a-mi, a-bo…*	'I, you…'	pron.	—
(22)	CC + 4	*/koroc/	*koroc* (CC + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1)[48]	CVCVC	—	atten.	No	*partí*	'give'	v.	—
(23)	CC + 6	*/kuŋ/	*kuŋ* (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	—	intens.	No	*yiñcí*	'be full'	v.	—

[48] A related form *koroc* ~ *ñoroc* was found in Geba (Geba) where it seems to be used only as an adjective (not as an ideophonic adverb).

	Continental UGC common ideophone							Associated element			Other CUGC equiv.
No.	**Distr. profile**	**Ideophone**	**Attested forms**	**SS**	**Etymon**	**SEM**	**RED**	**Form**	**Meaning**	**PoS**	
(24)	CC + 5	*/lot/	*lot* (CC + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1); *lɔt* (Geba)	CVC	—	intens.	No	*negá*	'refuse'	v.	—
(25)	CC + 6	*/nok/	*nok* (CC + Cacheu3 + Bissau1 + Bissau3); *nɔk* (Geba + Cacheu1 + Cacheu2)	CVC	Mandinka© (QM)	intens.	No	*pretu*	'be black'	qual.	—
(26)	CC + 4	*/nu(ŋ)/	*nu* (Geba + Cacheu1 + Cacheu3 + Bissau1); *nuŋ* (CC)	CVC	—	intens.	No	*kurpu-limpu*	'be naked'	qual	—
(27)	CC + 4	*/palaw/ ~ */paraw/	*palaw* (CC); *falaw* (Geba); *paraw* (Cacheu3 + Bissau2); *paw* (Cacheu1)	CVCVC	—	visual	No	*lamprá*	'shine'	v.	*pas ~ feles* (Geba)
(28)	CC + 4	*/parac/	*parac* (CC + Geba + Cacheu1 + Cacheu3); *parat* (Bissau1)	CVCVC	—	onom.	No	*kokó*	'defecate'	v.	—
(29)	CC + 4	*/pat/ ~ */put/	*pat* (CC + Cacheu3 + Bissau1); *fat* (Cacheu1); *parat* (Bissau1); *put* (CC + Geba)	CVC	Mandinka + Wolof©	intens.	No	*sapá*	'cut'	v.	—
(30)	CC + 6	*/pɔtɔk/	*pɔtɔk* (Cacheu1 + Cacheu2 + Cacheu3 + Bissau1); *potok* (Bissau3); *botok* (CC); *pɔcɔk* (Geba)	CVCVC	Mandinka©	intens.	No	*moli*	'be soft'	qual.	—
(31)	CC + 5	*/pɔtɔk/ ~ */pucuk/	*potok* (CC); *pɔtɔk* (Geba + Cacheu1); *putuk* (Cacheu2); *pɔcɔk* (Geba); *pucuk* (Cacheu3 + Bissau1); *pucak* (Bissau3)	CVCVC	Mandinka©	intens.	No	*susu*	'be dirty'	qual.	—
(32)	CC + 6	*/pus/	*pus* (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	—	intens.	No	*limpu*	'be clean'	qual.	—
(33)	CC + 5	*/rap/	*rap* (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3); *rip* (Cacheu3); *raŋ* (Bissau1)	CVC	Wolof©	intens.	No	*ficá*	'close'	v.	*vaŋ* (Bissau1)

Continental UGC common ideophone								Associated element			Other CUGC equiv.
No.	**Distr. profile**	**Ideophone**	**Attested forms**	**SS**	**Etymon**	**SEM**	**RED**	**Form**	**Meaning**	**PoS**	
(34)	CC + 6	*/sip/	*sip* (Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau2); *lip* (CC); *lip* (+ *fartá* 'be satisfied) (Cacheu3); *vip* (Geba)	CVC	—	intens.	No	*pisadu*	'be heavy'	qual.	—
(35)	CC + 4	*/taw/	*taw* (+ *malgós* 'be bitter') (CC); *taw* (+ *forti* 'be sour') (Geba + Cacheu1 + Cacheu2 + Cacheu3)	CVC	Mandinka + Wolof©	intens.	No	*malgós*	'be bitter'	qual.	*rok* (+ *malgós* 'be bitter') (Geba)
(36)	CC + 5	*/tip/	*tip* (CC + Geba + Cacheu1 + Cacheu3 + Bissau1); *tip* (+ *sukuru* 'be dark') (Bissau3); *sip* (Geba)	CVC	—	intens.	No	*pegá* (CC)/*duru* (others)	'thicken'	v.	—
(37)	CC + 6	*/wak/	*wak* (Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3); *wɛk* (CC)	CVC	—	intens.	No	*burmeju*	'be red'	qual.	—
(38)	CC + 6	*/wandaŋ/	*wandaŋ* (CC + Cacheu3 + Bissau1 + Bissau2); *wantaŋ* (+ *klaru* 'be clear(ed)') (Geba); *wandalaŋ* (Cacheu1 + Cacheu2)	CVCCVC	Mandinka© (QM)	intens. ('wide')	No	*yabrí*	'open (door)'	v.	—
(39)	CC + 6	*/wit/	*wit* (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	Mandinka©	intens.	No	*kintí*	'be hot'	qual.	—

Table 1.4: 'other distribution' type: 'CC vs. Guinea-Bissau' and 'no CC vs. Guinea-Bissau' subtypes.

Continental UGC ideophones									Associated element			Other CUGC equiv.
No.	**Distr. profile**	**CC**	**Guinea-Bissau**	**Attested forms**	**SS**	**Etymon**	**SEM**	**RED**	**Form**	**Meaning**	**PoS**	
(1)	CC vs. Guinea-Bissau (3)[49]	[lɐm]	*/waŋ/	*waŋ* (Geba + Cacheu3 + Bissau1)	CVC	—	intens.	No	*didiya*	'at noon'	adv.	—
(2)	CC vs. Guinea-Bissau (3)	[maj]	*/fis/	*fis* (Cacheu1 + Cacheu2); *fis* (+ *sai* 'get out') (Bissau3)	CVC	Mandinka ©	speech ('without a word')	No	*pasá*	'go by'	v.	—
(3)	CC vs. Guinea-Bissau (4)	[liŋ]	*/kaŋ/	*kaŋ* (Geba + Cacheu2 + Bissau1); *taŋ* (Cacheu3)	CVC	Wolof©	intens.	No	*tesu*	'be solid'	qual.	*ŋaw* (Geba)
(4)	no CC vs. Guinea-Bissau (2)	—	*/sip/	*sip* (Bissau3); *zip* (Geba)	CVC	—	intens.	No	*gros*	'be fat/ thick'	qual.	—
(5)	no CC vs. Guinea-Bissau (4)	—	*/far/	*far* (Geba + Cacheu1 + Cacheu2 + Bissau3)	CVC	Wolof©	intens.	No	*negá*	'refuse'	v.	—
(6)	no CC vs. Guinea-Bissau (5)	—	*/caw/[50]	*caw* (Geba + Cacheu2 + Bissau3); *cadaw* (Cacheu3 + Bissau1 + Bissau3)	CVC	Mandinka + Wolof©	intens. ('ripe fruit')	No	*burmeju*	'red'	qual.	—

[49] In Tables 1.4 and 1.5, the digits given in brackets in the column 'Distribution profile' correspond to the number of consultants who supported a given form.

[50] The form */caw/ has been retained as the base form for this ideophone against *cadaw*, which is almost certainly an extended (expressive) form of *caw* (see also */baw/ above).

Table 1.5: 'other distribution' type: '(CC + Guinea-Bissau) vs. Guinea-Bissau' subtype.

Continental UGC ideophones										Associated element			Other CUGC equiv.
(CC + Guinea-Bissau)				Guinea-Bissau		SS	Etymon	SEM	RED	Form	Meaning	PoS	
No.	**Distr. profile**	**CF**	**Attested forms**	**CF**	**Attested forms**								
(1)	(CC + Guinea-Bissau) (2) vs. Guinea-Bissau (3)	*/mut/	*mut* (CC + Geba)	*/mip/	*mip* (Cacheu3 + Bissau2 + Bissau3)	CVC	Wolof©	intens.	No	*sukuru*	'be dark'	qual.	—
(2)	(CC + Guinea-Bissau) (3) vs. Guinea-Bissau (3)	*/ɲos/	*ños* (CC + Cacheu3); *ñɔk* (Geba)	*/rek/	*rek* (Cacheu3 + Bissau1); *rek* (+ *justá* 'be enough') (Bissau3)	CVC	Wolof©	intens.	No	*cigá*	'be closely related (parents)'	v.	—
(3)	(CC + Guinea-Bissau) (2) vs. Guinea-Bissau (5)	*/mik/	*mik* (CC); *muruk* (Geba)	*/jem/	*yem* (Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	Mandinka + Wolof©	intens.	No	*kalá*	'keep quiet'	v.	*tik* (Cacheu2)

Appendix 2: Insular UGC ideophones

Table 2.1: Santiago ideophones.

	Santiago ideophone					Associated element			Continental UGC ideophone		
No.	**Ideophone**	**SS**	**Etymon**	**Semantic type**	**RED**	**Form**	**Meaning**	**PoS**	**Semantic Equiv.**[51]	**Cognate**	**Distr. type**
(1)	*álbu* [ˈalbu]	VCCV	Port. (§4.4.1)	intens.	No	*bránku*	‘white’	adj.	*/fandaŋ/	No	common core
(2)	*ban* [ˈbɛ̃]	CVC	—	intens.	No	*intxi*	‘fill’ ~ ‘be full’	v.	*/kuŋ/	No	common core
(3)	*bipu* [ˈbipu] ~ *vipu* [ˈvipu]	CVCV	—	intens.	No	*intxi*	‘fill’ ~ ‘be full’	v.	*/kuŋ/	No[52]	common core
(4)	*bódji-bódji* [ˈbɔdʒi ˈbɔdʒi]	CVCV-CVCV	Mandinka©	intens.	Yes	*gordu*	‘fat’	adj.	*/sip/	No	other
(5)	*bupu* [ˈbupu]	CVCV	—	onom.	No	*kai*	‘fall’	v.	*/**bup**/	Yes	common core
(6)	*fakati* [fɛˈkɛti]	CVCVCV	Mandinka + Wolof©	immdt.	No	*labánta*	‘stand up’	v.	*/**fa**(r)**kat**/ ~ */**fɛ**(r)**kɛt**/	Yes	common core
(7)	*fépu* [ˈfɛpu]	CVCV	Mandinka + Wolof© (QM)	intens.	No	*kumi/kába*	‘eat, finish’	v.	*/**fep**/	Yes	common core

[51] In the column ‘semantic equivalent’, the segments printed in bold justify the existence of an etymological relationship between insular and continental UGC ideophones, and they have lead us to write ‘Yes’ in the column ‘cognate’.

[52] A possible link could be posited with the continental form */sip/, which intensifies *pisadu* ‘be heavy’ (see Table 1.3), a lexical item semantically close to ‘be full’. Furthermore, we have noted a variant *vip* [vip] in Geba, which displays an even greater similarity with Capeverdean *vipu*. However, as *vip* was only attested once and as the semantic relationship between both associated elements (*pisadu* and *intxi*) is not absolutely straightforward, we have preferred not to consider these two ideophones as cognates within the scope of this study.

	Santiago ideophone					Associated element			Continental UGC ideophone		
No.	**Ideophone**	**SS**	**Etymon**	**Semantic type**	**RED**	**Form**	**Meaning**	**PoS**	**Semantic Equiv.**[51]	**Cognate**	**Distr. type**
(8)	*finu* [ˈfinu]	CVCV	Mandinka + Port.© (QM)	intens.	No	*prétu*	'black'	adj.	*/nok/	No	common core
(9)	*futi* [ˈfuti]	CVCV	Mandinka + Wolof©	immdt.	No	*sai*	'get out'	v.	*/**fut**/	Yes	common core
(10)	*fututú* [futuˈtu]	CVCVCV	—	intens.	Yes	*txera*	'smell' (intr.)	v.	[**fututu**t]	Yes	CC-only
(11)	*góbu-góbu* [ˈgɔbu ˈgɔbu]	CVCV-CVCV	Mandinka© (QM)	intens.	Yes	*bedju*	'old'	adj.	*/kokorot/	No	minority
(12)	*ien* [ˈjẽ] ~ *ian* [ˈjẽ]	CVC	Mandinka + Wolof©	intens.	No	*kála*	'be silent'	v.	*/**jem**/	Yes	other
(13)	*iondon* [jõˈdõ]	CVCCVC	Wolof©	visual ('desert')	No	*fika*	'be'	v.	*/w**and**al**aŋ**/	Yes	minority
(14)	*kati* [ˈkɛti]	CVCV	—	intens.	No	*kebra*	'break'	v.	*/**kat**/	Yes	common core
(15)	*kóti-kóti* [ˈkɔti ˈkɔti]	CVCV-CVCV	Mandinka© (QM)	intens.	Yes	*bedju*	'old'	adj.	*/kokorot/	No	minority
(16)	*lápu* [ˈlapu]	CVCV	Wolof©	intens. ('tight')	No	*xinta*	'sit'	v.	*/b**ap**/ ~ [v**ap**] (Bissau1)	Yes	minority
(17)	*mápu* [ˈmapu]	CVCV	—	intens.	No	*pánha / pega / da*	'seize, take'	v.	[**mɐp**]	Yes	CC-only
(18)	*moku* [ˈmoku]	CVCV	Wolof© (QM)	intens.	No	*pila/kebra*	'crush'	v.	*/ˈ**moku**/ (Bissau3 + CC)	Yes	NA
(19)	*muku-muku* [ˈmuku ˈmuku]	CVCV-CVCV	Mandinka + Wolof©	onom. ('silent')	Yes	*fika*	'be'	v.	[**mik**] ~ [**mu**r**uk**]	Maybe	other
(20)	*mus-mus* [ˈmus ˈmus]	CVC-CVC	Wolof©	intens.	Yes	*perdi*	'disappear'	v.	[**mes**]	Yes	CC-only

	Santiago ideophone					Associated element			Continental UGC ideophone		
No.	**Ideophone**	**SS**	**Etymon**	**Semantic type**	**RED**	**Form**	**Meaning**	**PoS**	**Semantic Equiv.**[51]	**Cognate**	**Distr. type**
(21)	*(nun)priti* [nũˈpriti]	CCVCV	—	intens. (‘stark naked’)	No	*fika*	‘be’	v.	*/nu(ŋ)/	No	common core
(22)	*pan-pan* [ˈpɐ̃ ˈpɐ̃]	CVC-CVC	—	intens.	Yes	*ségu*	‘blind’	adj.	—	Zero	NA
(23)	*plátxi-plátxi* [ˈplatʃi ˈplatʃi]	CCVCV-CCVCV	—	intens.	Yes	*gordu*	‘fat’	adj.	*/sip/	No	other
(24)	*pran-pran* [ˈprɐ̃ ˈprɐ̃]	CCV-CCV	—	speech (‘clearly’)	Yes	*fla/papia*	‘say, talk’	v.	[pɐ̝jɐ̝j]	No	CC-only
(25)	*pratxi* [ˈprɐtʃi]	CCVCV	—	onom.	No	*fasi kokó*	‘defecate’	VP	*/**p**a**rac**/	Yes	common core
(26)	*prikiti* [priˈkiti]	CCVCVCV	Mandinka + Wolof©	immdt.	No	*labánta*	‘stand up’	v.	*/**pirkit**/	Yes	minority
(27)	*pufu* [ˈpufu]	CVCV	—	intens.	No	*fra*	‘make a hole, pierce’	v.	*/**puf**/	Yes	minority
(28)	*saki* [ˈsɐki] ~ *siki* [ˈsiki]	CVCV	—	intens.	No	*sápa*	‘get cut’	v.	*/pat/ ~ */put/	No	common core
(29)	*tran* [ˈtrɐ̃]	CCVC	—	intens.	No	*sakédu*	‘stand’	v.	*/caŋ/ ~ */kaŋ/ ~ [**taŋ**] (Bissau1)	Maybe	other
(30)	*txápu* [ˈtʃapu]	CVCV	Mandinka©	intens.	No	*pánha / pega / da*	‘seize, take’	v.	*/**cap**/	Yes	common core
(31)	*txulupu* [tʃuˈlupu] ~ *txulufu* [tʃuˈlufu] ~ *txufun* [tʃuˈfũ]	CVCVCV ~ CVCVCV ~ CVCVC	—	onom. (‘splash’)	No	*da*	‘give’	v.	*/**culup**/	Yes	common core

Santiago ideophone						Associated element			Continental UGC ideophone		
No.	**Ideophone**	**SS**	**Etymon**	**Semantic type**	**RED**	**Form**	**Meaning**	**PoS**	**Semantic Equiv.**[51]	**Cognate**	**Distr. type**
(32)	*uandan* [wɛ̃ˈdɛ̃] ~ *iandan* [jɛ̃ˈdɛ̃]	CVCCVC	Mandinka© (QM)	intens.	No	*abri*	‘open’	v.	*/**wandaŋ**/	Yes	common core
(33)	*uátchi* [ˈwatʃi]	CVCV	Mandinka©	intens.	No	*fendi*	‘split open’	v.	*/**wac**/	Yes	minority
(34)	*uís* [ˈwis]	CVC	—	intens.	No	*lolu*	‘slip’	v.	[**fit**] (CC)	No	NA

Appendix 3: Etymology of some UGC ideophones

Table 3.1: Origins of those African-derived ideophones for which a possible etymon was identified.

UGC	Proto-CUPC form */// or Santiago [] form	Origin	Table	Distribution category	Etymon
Continental	*/caw/	Mandinka + Wolof	1.4	other (no CC vs. Guinea-Bissau)	Mandinka ***cáw*** (IDEO) 'very [red], intensely [red]' (Creissels 2012: 33); Wolof ***coy*** [cɔj] (IDEO) 'very [red]' (Diouf 2003: 85)
Continental	[colop]	Mandinka	1.1	CC-only	Mandinka ***cúrúm*** (IDEO) '[jump] quickly [down]' (Creissels 2012: 34)
Continental	*/corot/	Mandinka	1.3	common core	Mandinka ***córóti*** (v.) 'drip, be tiny' (Creissels 2012: 34)
Continental	*/far/	Wolof	1.4	other (no CC vs. Guinea-Bissau)	Wolof ***fétt*** [fe:ttə] (IDEO) '[refuse] adamantly, strongly [refuse]' (Diouf 2003: 126)
Continental	*/fas/	Wolof	1.2	minority	Wolof ***fés*** [fes] (v.) 'be visible/conspicuous' (Diouf 2003: 126)
Continental	*/fis/	Mandinka	1.4	other (CC vs. Guinea-Bissau)	Mandinka ***fíw*** (IDEO) '[go by] quickly' (Creissels 2012: 70)
Continental	*/fit/	Mandinka	1.3	common core	Mandinka ***fít*** (IDEO) '[go by] quickly' (Creissels 2012: 69); ***fálás*** (IDEO) '[go by] quickly' (Creissels 2012: 58)
Continental	*/fitfat/	Mandinka	1.3	common core	Mandinka (IDEO) ***páfát*** ~ ***pápát*** '[flee] quickly' (Creissels 2012: 200)
Continental	[fof]	Mandinka	1.1	CC-only	Mandinka ***yéf*** (IDEO) 'very [light]' (Creissels 2012: 271)
Continental	*/fuf/ ~ */fus/	Wolof	1.2	minority	Wolof ***fuuf*** [fu:f] (adv.) 'much [more]' (Diouf 2003: 132)
Continental	*/jem/	Mandinka	1.3	common core	Mandinka ***yóm*** [jóm] (IDEO) 'very [cold]' (Creissels 2012: 273)
Continental	*/jɔp/	Wolof	1.3	common core	Wolof ***xepp*** [xɛppə] (IDEO) 'completely [wet]' (Diouf 2003: 387; Fal et al. 1990: 256)
Continental	[kamaɟ]	Wolof	1.1	CC-only	Wolof ***kamaj*** [kamac] (IDEO) '[be] suddenly [switched off]' (Diouf 2003: 181)
Continental	*/kaŋ/	Wolof	1.3 +1.4	common core + other (CC vs. Guinea-Bissau)	Wolof ***këŋ*** [kəŋ] (IDEO) 'very [hard/tough]' (Diouf 2003: 185)

UGC	Proto-CUPC form */// or Santiago [] form	Origin	Table	Distribution category	Etymon
Continental	[koŋ]	Wolof	1.1	CC-only	Wolof ***koŋ*** [kɔŋ] (IDEO) 'very [dry]' (Diouf 2003: 188)
Continental	*/las/	Wolof	1.2	minority	Wolof ***làcc*** [la:ccə] (v.) 'be exhausted' (Diouf 2003: 193)
Continental	*/mut/	Wolof	1.5	other ((CC + Guinea-Bissau) vs. Guinea-Bissau)	Wolof ***muut*** [mu:t] (v.) 'keep calm, stand still' (Diouf 2003: 231)
Continental	*/nok/	Mandinka	1.3	common core	Mandinka ***nót*** (IDEO) 'very (black and dirty)' (Creissels 2012: 188)
Continental	*/pat/ ~ */put/	Mandinka +Wolof	1.3	common core	Mandinka ***pát*** (IDEO) '[cut] quickly' (Creissels 2012: 201); ***pókót*** '[cut] silently' (Creissels 2012: 203); Wolof ***pàcc*** (IDEO) '[cut] into two equal parts' (Fal et al. 1990: 256)
Continental	*/pɔtɔk/	Mandinka	1.3	common core	Mandinka ***páták*** (IDEO) 'very [wet]' (Creissels 2012: 201); *pòtó* (n.) 'clay, mud, silt' (Creissels 2012: 203)
Continental	*/pɔtɔk/ ~ */pucuk/	Mandinka	1.3	common core	Mandinka ***hót*** (IDEO) 'very [black and dirty]' (Creissels 2012: 82)
Continental	[pul]	Wolof	1.1	CC-only	Wolof ***pëll*** [pəllə] (IDEO) '[get out] quickly' (Diouf 2003: 277-278; Fal et al. 1990: 169)
Continental	*/rap/	Wolof	1.3	common core	Wolof ***ràpp*** [ra:ppə] (IDEO) '[close] tightly' (Diouf 2003: 285)
Continental	*/rek/	Wolof	1.5	other ((CC + Guinea-Bissau) vs. Guinea-Bissau)	Wolof ***rekk*** [rɛ:kkə] (adv.) 'only' (Diouf 2003: 289); ***rəkk*** [rəkkə] ~ ***rikk*** [rikkə] (IDEO) 'just [beside], right [next to]' (Diouf 2003: 289, 291)
Continental	*/taw/	Mandinka +Wolof	1.3	common core	Mandinka ***láw*** (IDEO) 'very [bitter]' (Creissels 2012: 156); Wolof ***toll*** [tɔllə] (IDEO) 'very [bitter]' (Diouf 2003: 347)
Continental	*/tep/	Mandinka	1.2	minority	Mandinka ***tép*** (IDEO) 'completely [full], [filled] to the brim' (Diouf 2003: 58)
Continental	[was]	Mandinka +Wolof	1.1	CC-only	Mandinka ***wásák*** (IDEO) '[pour/spill] all over', expresses the idea of 'dispersal' or 'scattering' (Creissels 2012: 264)

UGC	Proto-CUPC form *// or Santiago [] form	Origin	Table	Distribution category	Etymon
					and/or Wolof ***wesar*** [wɛsar] (v.) 'be scattered' (Diouf 2003: 368)
Continental	*/wit/	Mandinka	1.3	common core	Mandinka ***wíj*** [wiɟ] (IDEO) 'very [hot]' (Creissels 2012: 265)
Continental +Insular	*/fep/ + [ˈfɛpu]	Mandinka +Wolof	1.3 +2.1	common core	Mandinka ***féw*** (IDEO) '[finish] completely' (Creissels et al. 1982: 47, Rougé 2004: 308) and/or Wolof ***–épp*** (adv.) 'completely'; ***fépp*** [**fe:pp**ə] (adv.) 'complete, every(where)' (Diouf 2003: 125)
Continental +Insular	*/wandaŋ/ + [wɛ̃ˈdɛ̃] ~ [jɛ̃ˈdɛ̃]	Mandinka	1.3 +2.1	common core	Mandinka ***wáráŋ*** (IDEO) '[swing] wide open' (Creissels 2012: 264)
Continental +Insular	*/ˈmoku/ + [ˈmoku]	Wolof	2.1	NA	Wolof *mokk* [**mɔ:kk**ə] (v.) 'be ground/crushed' (Dieng 1985: 247; Diouf 2018: 226)
Continental +Insular	*/fut/ + [ˈfuti]	Mandinka +Wolof	1.3 +2.1	common core	Mandinka ***farafat*** (IDEO) '[come out] suddenly' (Creissels 2012: 61) and/or Wolof ***poset*** (IDEO) '(come out) suddenly' (Fal et al. 1990: 171)
Continental +Insular	*/jem/ + [ˈjẽ] ~ [ˈjɛ̃]	Mandinka +Wolof	1.5 +2.1	other ((CC + Guinea-Bissau) vs. Guinea-Bissau)	Mandinka ***yéréŋ*** (IDEO) 'very [quiet]' and/or Wolof *remm* [r**ɛ:m**mə] (IDEO) 'quiet(ly) (sea)' (Diouf 2003: 289); *xerem* [xɛr**ɛm**] 'completely [quiet]' (Diouf 2003: 387)
Continental +Insular	*/yambalaŋ/ (CC) ~ */wandalaŋ/ (Guinea-Bissau) + [jõˈdõ] (Santiago) + [ẽˈdẽ] (Fogo)	Wolof	1.2 +2.1	minority	Wolof ***yàmbalaŋ*** [ja:mbalaŋ] (IDEO) 'wide [open], extensive, large' (Diouf 2003: 398). The CC is obviously derived from Wolof. The other forms may have been influenced by Mandinka *wáráŋ* 'wide [open]' which is the most plausible etymon of continental */wandaŋ/ and insular [wɛ̃ˈdɛ̃] ~ [jɛ̃ˈdɛ̃] 'wide (open)' (Quint & Moreira Tavares 2019: 158). Some Santiago speakers consider that [wɛ̃ˈdɛ̃] ~ [jɛ̃ˈdɛ̃] can be used indifferently in lieu of [jõˈdõ] while others maintain a distinction [jõˈdõ] to refer to a 'quiet, empty place' vs. [wɛ̃ˈdɛ̃] to intensify the 'opening of a door'. We have considered here that the two roots are separated, which is probably historically true.

UGC	Proto-CUPC form *// or Santiago [] form	Origin	Table	Distribution category	Etymon
Continental + Insular	*/bap/ + [ˈlapu]	Wolof	1.2 + 2.1	minority	Wolof ***mbàpp*** [mbaːppə] (IDEO) (Diouf 2003: 217) ~ ***wàpp*** (IDEO) (Fal et al. 1990: 239) '[sit down] heavily'
Continental + Insular	*/mik/ + [ˈmuku ˈmuku]	Mandinka + Wolof	1.2 + 2.1	other ((CC + Guinea-Bissau) vs. Guinea-Bissau)	Mandinka ***méj*** (IDEO) 'very [quiet]' (Creissels 2012: 170) and/or Wolof *miig* [**miːk**] (itj.) 'hush, shh' (Diouf 2003: 225); ***tɛkk*** [tɛːkkə] 'very (quiet)'
Continental + Insular	[mes] + [ˈmus ˈmus]	Wolof	1.1 + 2.1	CC-only	Wolof ***mes*** [mɛs] (IDEO) '[disappear] suddenly' (Diouf 2003: 225; Fal et al. 1990: 131)
Continental + Insular	*/pirkit/ + [priˈkiti]	Mandinka	1.2 + 2.1	minority	Mandinka ***kíríbít*** (IDEO) '[stand up] suddenly' (Creissels 2012: 127)
Continental + Insular	*/cap/ + [ˈtʃapu]	Mandinka	1.3 + 2.1	common core	Mandinka ***cápí*** (v.) 'take, seize' (Creissels 2012: 33); ***cás*** (IDEO) '[catch] brutally [in the air]' (Creissels 2012: 33); ***táp*** (IDEO) '[seize] quickly' (Creissels 2012: 244)
Continental + Insular	*/wac/ + [ˈwatʃi]	Mandinka	1.2 + 2.1	minority	Mandinka ***wác*** ~ ***wáj*** (IDEO) 'rip, brutally [torn]' (Creissels 2012: 263)
Continental + Insular	*/fa(r)kat/ ~ */fɛ(r)kɛt/ + [fɛˈkɛti]	Mandinka + Wolof	1.3	common core	Mandinka *có**kót*** (IDEO) '[stand up] at once' (Creissels 2012: 33); ***táráfát*** (IDEO) '[stand up] suddenly' (Creissels 2012: 245) and/or Wolof *ño**kket*** [ɲɔːkkɛt] (IDEO) '[stand up] at once' (Diouf 2003: 269); ***fojjet*** [fɔːɟɟɛt] (IDEO) '[stand up] immediately' (Diouf 2003: 128)
Insular	[ˈbɔdʒi ˈbɔdʒi]	Mandinka	2.1	common core	Mandinka ***bòjiboji*** (v.) 'be too ripe [fruit], too oily [dish]' (Creissels et al. 1982: 16)
Insular	[ˈfinu]	Mandinka + Port.	2.1	common core	Mandinka ***fíŋ*** (IDEO) 'black' (Creissels 2012: 67; Rougé 2004: 309) and/or Port. *fino* [ˈ**finu**] (adj.) 'fine, refined'
Insular	[ˈgɔbu ˈgɔbu]	Mandinka	2.1	minority	Mandinka ***kób*** (IDEO) 'very (old)' (Creissels 2012: 128)
Insular	[ˈkɔti ˈkɔti]	Mandinka	2.1	minority	Mandinka ***kòtó/kòtóo*** (qual./n.) '(be) old, elder' (Creissels 2012: 137)

Abbreviations (only used in the Appendices)

© = African etymon given in Appendix 3
adj. = adjective
atten. = attenuation
CF = common form
CUGC = Continental Upper Guinea Creole
Distr. = distribution
equiv. = equivalent
immdt. = immediate
itj. = interjection
intens. = intensifying/intensive
n. = noun
onom. = onomatopoeia (= sound ideophone)
PoS = Part of Speech
pron. = pronoun
(QM) = published in Quint & Moreira Tavares (2019: 157-158)
qual. = qualifier
SEM = semantic type
SS = syllable structure
v. = verb
VP = verb phrase