

# Initial vowel agglutination in the Gulf of Guinea creoles<sup>1</sup>

Tjerk Hagemeijer

## 1. Introduction

Reinterpretation of morpheme boundaries is a well-attested phenomenon in contact linguistics and language-internal diachronic change. Examples of agglutination have been noted in a wide array of creole languages (e.g. Holm 1988: 97; Parkvall 2000: 81-3) and especially in French-based creoles (e.g. Baker 1984, Grant 1995). This paper focuses on the Gulf of Guinea creoles (GGCs), where a number of etymologically consonant-initial words in the lexifier language, Portuguese, exhibit an agglutinated vowel lacking a morphological function. This property is particularly common in Lung'ie (Principense Creole). My aim is to answer the following interrelated questions:

- (i) Is there evidence for diachronic layering of agglutination in the GGCs?
- (ii) What are the workings that underlie agglutination in the GGCs?
- (iii) What are the origins of agglutination in the GGCs?

In a nutshell, the four GGCs are the 16<sup>th</sup> century offspring of a proto-GGC (e.g. Ferraz 1979, Hagemeijer 1999, Schang 2000, Hagemeijer & Parkvall 2001), a contact variety spoken on the island of São Tomé that resulted from contact between Portuguese and African languages, in particular from the Niger Delta and the Congo/Angola area. The proto-GGC evolved into what is presently Santome (São-Tomense). Lung'ie is the historical result of a variety or varieties of the proto-GGC whose speakers settled on the island of Príncipe, located to the north of São Tomé. The settling of creole speakers from São Tomé on the island of Annobón, located to the south of São Tomé, created an isolated variety now known as Fa d'Ambu (Annobonese). Finally, Ngola (Angolar), spoken on S. Tomé, is arguably the historical outcome of a society of runaway slaves (Lorenzino 1998).

This paper has the following outline. Section 2 discusses the agglutination data in Appendix 1. In section 3 I will provide an overview of how agglutination in the GGCs has

---

<sup>1</sup> I am very grateful to Norval Smith for comments on draft versions of this paper. I would also like to thank the audience of the ACBPLe 2005 Orléans meeting, where I first presented this paper, especially Thomas Klein. I want to express especial gratitude to Philippe Maurer and Armando Zamora for comments and for letting me have access to their unpublished work on Lung'ie and Fa d'Ambô, respectively. Finally, a word of thank to Marike Post and Ota Ogie for helping me out with the Fa d'Ambô and Édó data, respectively. After sending a final draft version to the editors, it came to my knowledge that John Ladhams was awaiting the publication of a paper called "Article Agglutination and the African Contribution to the Portuguese-based Creoles" (see references), which deals with the same topic but in a distinct fashion.

been treated in previous literature. Section 4 considers the role of the Portuguese article system with respect to prosthesis. Section 5 is concerned with vowel assimilation and harmony processes that play a role in agglutination. In section 6 I will address the putative role of the different strata that contributed to the formation of the GGCs with regard to the topic under consideration.

## 2. Overview of the data

A full list of agglutinated items in the GGCs and the sources they were drawn from can be found in Appendix 1, which shows that agglutination is an outstanding feature of Lung'ie. Note that this does not pretend to be an exhaustive list of items.

Since diffusion of the GGCs took place from the proto-GGC that evolved into Santome, I will first consider to what extent the agglutinations found in this language are also found in the other GGCs. The results are listed in Table 1.<sup>2</sup>

Table 1. Instances of agglutination in Santome with matches in the other GGCs.<sup>3</sup>

Santome	Lung'ie	Fa d'Ambu	Ngola	Portuguese etymology	Meaning
<b>afe</b>	<b>afe</b>	fe	fe	<i>fé (fem.)</i>	'faith'
<b>apa</b>	<b>apa</b>		<b>anpa</b>	<i>pá (fem.)</i>	'spade, shovel'
<b>ase</b>	<b>(gêêza)-ase</b>	gêza	ngeedha	<i>sé (fem.)</i>	'cathedral'
<b>aglasa</b>	nomi	galasa, nomi	nomi	<i>graça (fem.)</i>	'name'
<b>ope</b>	<b>ope</b>	<b>ope</b>	<b>ope</b>	<i>pé (masc.)</i>	'foot, leg'
<b>opo</b>	<b>opo</b>	<b>opo</b>	<b>ompo</b>	<i>pó (masc.)</i>	'dust'
<b>ose</b>	<b>ose</b>	<b>ôsê</b>	<b>onthe</b>	<i>céu (masc.)</i>	'sky, heaven'
<b>ono</b>	<b>ono</b>	<b>ono</b>		<i>nó (masc.)</i>	'knot'
<b>odo</b>	<b>odo</b>		<b>odo~oro</b>	<i>dó (masc.)</i>	'mourning'
<b>ôlhô<sup>4</sup></b>	<b>ôryô</b>	lubela	awa	<i>rio (masc.)</i>	'river'
<b>oali</b>	<b>uari</b>	oventu	ventu	<i>ar (masc.)</i>	'air'
<b>omali</b>	<b>umwe(n)</b>	<b>omal</b>	mionga	<i>mar (masc.)</i>	'sea'
<b>uku</b>	<b>uku</b>		ndagu	<i>cu (masc.)</i>	'arse'
<b>unu</b>	<b>unu</b>	dodyi		<i>nu (masc.)</i>	'naked'

First, it can be observed that in most cases agglutinated items in Santome have a corresponding form in Lung'ie and to a certain extent in Fa d'Ambu and Angolar as well. It should be noted that some items in Ngola, such as *mionga* and *nvumbu*, have a Bantu origin,

<sup>2</sup> Note that for writing purposes closed mid-vowels are written *ê* and *ô* and open mid-vowels *e* and *o*.

<sup>3</sup> I did not include the item *arê~alê* 'king', which is common to the four GGCs, because it is derived from Old Portuguese expression *el-rei* 'the king' rather than form *rei*.

<sup>4</sup> This item occurs in Negreiros (1895) but is absent in contemporary speech.

which is expected in the particular scenario of relexification proposed by Lorenzino (1998). The open cells in Table 1 simply reflect the lack of available data. Except for *aglasa* ‘name’, it follows that these shared items involve essentially etymologically monosyllabic words with an agglutinated [a], [ɔ], [o], or [u]. It can be concluded that these forms reflect an early diachronic stage and were lexicalized in the proto-GGC, prior to the spatio-temporal diffusion into four different creoles sketched in section 1.

Fa d’Ambu, on the other hand, exhibits several cases of agglutination that are shared with Lung’ie but are not found in Santome nor Ngola, as illustrated in Table 2.

Table 2. Instances of agglutination common to Lung’ie and Fa d’Ambu.

Fa d’Ambu	Lung’ie	Santome	Ngola	Portuguese etymology	Meaning
<b>ôzôyô</b>	<b>ôzê (n)~ôzên~uzên</b>	zê	m’puna~puna	<i>joelho</i>	‘knee’
<b>onfenu</b>	<b>unfenu</b>	nfenu	nfenu	<i>inferno</i>	‘hell’
<b>ôlatu</b>	<b>uratu</b>	latu	latu	<i>rato</i>	‘rat, mouse’
<b>olemu</b>	<b>uremu</b>	lemu	lemu~lêmu	<i>remo</i>	‘oar’
<b>oventu</b>	<b>uventu</b>	ventu	ventu	<i>vento</i>	‘wind’
<b>onfelu</b>	<b>ufew</b>	felu	felu~fêlu	<i>ferro</i>	‘iron’
<b>ampan</b>	<b>umpan</b>	mpon	pon	<i>pão</i>	‘bread’
<b>ônũa, ônũaya</b>	<b>unwan</b>	nũa	mbêi~mbêzi~ mêzi	<i>lua</i>	‘moon’
<b>ôman</b>	<b>uman</b>	mon	mo~mon	<i>mão</i>	‘hand, lower arm’
<b>ôpa</b>	<b>upa</b>	po	po	<i>pau</i>	‘stick, tree’
<b>ôbôyô</b>	<b>ubwê</b>	bwê	buê	<i>bói</i>	‘cow’
<b>ôtesa</b>	tesa	tesa		<i>testa</i>	‘forehead’

The main difference is that in these cases Lung’ie exhibits almost exclusively prosthetic [u] and Fa d’Ambu typically an agglutinated [o] or [ɔ]. Moreover, I only found one agglutinated item in Fa d’Ambu that is absent in Lung’ie, namely *ôtesa* ‘forehead’. The structure of the Lung’ie item, with the consonant cluster, however, suggests that it has been (re)introduced in more recent times. It follows from Table 2 that at this point agglutination is no longer restricted to items whose etymology is monosyllabic. This raises the question as to why these agglutinated forms are not found in Santome as well, under the assumption that they spread from the proto-GGC, especially given the lack of historical evidence for direct contact between Lung’ie and Fa d’Ambu. I will return to this issue in section 6.3.

The numerous remaining cases involving agglutination listed in Appendix 1, i.e. the cases not listed in Table 1 and 2, are exclusive to Lung’ie, encompassing [u] and [i]-agglutination. Note that the latter type is not attested in any other GGC.

In sum, in the light of the data above, initial vowel agglutination first affected items with a monosyllabic etymology but spread to items with a disyllabic etymology. It is therefore suggestive that in the case of *omali* ‘sea’ (from Ptg. *mar*) and *oali* ‘air’ (from Ptg. *ar*) agglutination preceded resyllabification: *mar* > \**omar* > \**omari* > *omali*.

### 3. Previous literature on agglutination in the GGCs

Several scholars mention or seek to explain the existence of agglutination (Barrena 1957; Ferraz 1979, 1984; De Granda 1986; Parkvall 2000; Schang 2000).

In his work on Santome, Ferraz states that Portuguese monosyllabic nouns are sometimes integrated in the creole by agglutination of a non-nasal vowel, which he suggests is derived from the Portuguese definite article. He further suggests some sort of relation between the agglutinated forms and massive imports of Bantu-speaking slaves, but does not elaborate on this.

In a similar vein, De Granda (1986: 118) argues that prosthetic vowels in Fa d’Ambu are formally identifiable with the Portuguese singular definite *o* for masculine and *a* for feminine (cf. section 4). However, the two cases of [a]-agglutination listed by the author are attached to two items with a masculine etymology, *ampan* ‘bread’ (from (*o*) *pão*) and *alê* ‘king’ (from *el-rei*, see footnote 3). De Granda provides a number of examples of etymologically monosyllabic items that exhibit agglutination and argues that the purpose of this process is to produce disyllabic items. However, this claim cannot be upheld given the number of agglutinations to items in Fa d’Ambu (v. Appendix 1) whose etymology is disyllabic.

Parkvall (2000) introduces a different approach to agglutination by suggesting that the Edoid cluster is a putative substratum, since nouns are obligatorily vowel-initial<sup>5</sup> and Edoid influence on the GGCs is well attested. Following Ferraz’ (1979) claim that all verbs are consonant-initial in Santome, Parkvall sees in this a further link with the equally consonant-initial nature of verbs in this cluster.

Schang (2000) provides the most exhaustive discussion so far on prosthetic vowels in the GGCs, adopting the Bantu perspective summarily proposed in Ferraz. First, he considers that the influence from Bantu noun classifiers cannot be responsible for [i] and [u]-agglutination because a semantic overlap with the respective noun classes cannot be

---

<sup>5</sup> In this respect, note that Portuguese (or English) loanwords are/were integrated in Edo with an initial vowel (a, e, ɛ, i) in order to comply with the rule that all nouns should be vowel initial. All the oral vowels in Edo may serve as prefixes and their choice is generally not predictable (Agheyisi 1990: 32). A few examples of Portuguese nouns in Edo are: *efêrhinye* ‘a dish from unsifted cassava’ (Ptg. *farinha*), *ekuye* ‘spoon’ (Ptg. *colher*), *esara* ‘saw’ (Ptg. *serra*), *etuheru* ‘scissors’ (Ptg. *tesoura*), *epipa* ‘keg, barrel’ (Ptg. *pipa*), *ibata* ‘shoe, boot’ (Ptg. *bota*), *isâhê* ‘key’ (Ptg. *chave*), *itaba* ‘tobacco’ (Ptg. *tabaco*), etc. (Melzian 1937).

recovered.<sup>6</sup> Instead, hypothesis to explain [u]-agglutination in Lung'ie is based on the assumption that Portuguese lexicon massively relexifies Kimbundu lexicon, and that therefore Portuguese vocabulary was arguably incorporated into noun class 9 (loanwords) and additionally prefixed with Kimbundu's definite article [o] Schang further suggests that the similarities between Portuguese definite article [u] and Kimbundu [o] may have conflated into the pattern observed in Lung'ie. As a piece of confirmatory evidence for the relevance of the Bantu substrate, he argues that [i] and [u]-agglutination resemble [o], [ɔ] and [i]-prefixation of Portuguese lexicon produced by Angolan Tonga's<sup>7</sup>, who are essentially the descendants of contract labourers that arrived during the coffee and cacao boom in the 19<sup>th</sup> and 20<sup>th</sup> century, and in that sense unrelated to the creolization process that took place in the 16<sup>th</sup> century.

There are, however, several arguments that cast serious doubts on the alleged influence of Bantu on prosthesis. Foremost, it is relatively uncontroversial that Lung'ie exhibits a number of properties that clearly reflect a stronger impact from a Niger delta substrate than the other GGCs (section 6.3). And since these creoles exhibit a stronger Bantu impact, especially at the level of lexicon and phonology, one would expect more pervasive agglutination under the Bantu hypothesis. Ngola, in particular, exhibits an exceptionally high percentage of Kimbundu items (Maurer 1992, 1995, Lorenzino 1998), but yet shows the lowest degree of agglutination.<sup>8</sup> The lack of correlation between agglutination and a Bantu substrate is underscored by cross-linguistic evidence. Colombian Palenquero creole, for example, is heavily influenced by a Kikongo substrate (e.g. Schwegler 2006) but exhibits no agglutination of the type under discussion (Schwegler, p.c.).<sup>9</sup>

Furthermore, the morpho-syntax of the GGCs shows a greater deal of similarities with its Nigerian substrate than with its Bantu substrate (Hagemeyer 1999, 2005b; Hagemeyer & Parkvall 2001). The claim that Portuguese lexicon massively replaced Kimbundu lexicon in SLA is also unwarranted because the evidence suggest that Kimbundu played essentially an adstratal role at a stage when the creole(s) had already started crystallizing. The fact that the Tonga Portuguese samples do suggest some morpho-semantic transfer of a Bantu noun class system therefore sharply contrasts with the findings for Lung'ie, where this relation cannot be established.

---

<sup>6</sup> Although agglutination in the GGCs is not restricted to specific semantic classes, it should be noted that some semantic fields are overrepresented, which is especially the case for body-related items and mass nouns. The latter type, for instance, is clearly overrepresented in the small group of [i]-agglutination.

<sup>7</sup> On Tonga Portuguese, see Rougé (1992) and Baxter (2004).

<sup>8</sup> Note, however, that Maurer (1995: 26) proposes that nasal sonorants in Angolar can be interpreted as class-prefixes.

<sup>9</sup> Norval Smith [p.c.] brings to my attention that Palenquero does exhibit a single case of functional nominal prefixation directly reflecting the Bantu class prefix *ma-*.

Finally, if conflation of Bantu [o] and Portuguese [u], which both have a semantic function (definiteness), were responsible for the agglutinated forms, one would actually expect the agglutinated items to exhibit a semantic function as well, contrary to fact.

In sum, contrary to previous claims, the available evidence does not support a decisive role of Bantu in the agglutination patterns found in the GGCs in general and in Lung'ie in particular. On the other hand, the highlighted importance of Portuguese gender for the agglutinated forms appears to be motivated and will be addressed in the next section.

#### 4. (Mis)calquing on gender and number

In other creoles, especially French-based creoles, cases of agglutination typically involve the lexifier's determiner. In Portuguese, the lexifier language of the GGCs, definite articles bear gender and number features, as shown in Table 3.

Table 3. Number and gender marking in Portuguese.

	<b>Masculine</b>	<b>Feminine</b>
<b>Singular</b>	o [u]	a [ɐ]
<b>Plural</b>	os [uʃ]	as [ɐʃ]

Under the hypothesis that [o], [ɔ] and [u]-agglutination are potentially related to instances of Portuguese singular masculine *o*, and [a] to instances of Portuguese singular feminine *a*, it follows that agglutinations in the oldest diachronic layer, which corresponds to the items listed in Table 1, are, without exception, correctly calqued on the superstrate's gender distinction.

In the diachronic stage where similar agglutinations are found in Lung'ie and Fa d'Ambu (see Table 2), but crucially not in Santome and Ngola, several gender mismatches can be found, namely *unwan/ônũa* 'moon' (Ptg. *a lua*) and *uman/ôman* 'hand, lower arm' (Ptg. *a mão*) and *ampan* 'bread' (Ptg. *o pão*). These mismatches are also attested in agglutinated items that are exclusive to Lung'ie (see Appendix 1), for instance *ubuka* 'mouth' (Ptg. *a boca*), *ufaka* 'knife' (Ptg. *a faca*) and *usuva* 'rain' (Ptg. *a chuva*). Furthermore, the [i]-agglutination patterns exclusive to Lung'ie are unrelated to Portuguese gender distinctions.

Calquing on Portuguese is not limited to the relation between the agglutinated vowel and Portuguese gender, but also applies to agglutination of Portuguese plural forms. This follows from the few cases where creole items reanalyzed a Portuguese plural article. These forms are listed here in Table 4.

Table 4. The role of Portuguese number-marking in agglutination.

Form	Portuguese etymology	Meaning
<i>zonda</i> (LU)	<i>as ondas</i>	wave(s)
<i>zawa</i> (ST) <sup>10</sup>	<i>as águas</i>	urine (lit. the waters)
<i>zalima</i> (LU, ST), <i>zalma</i> (FA)	<i>as almas</i>	souls

In these cases, unvoiced fricative [ʃ] in the Portuguese plural article *as* becomes voiced [z] in intervocalic environments, namely *as ondas* [ɐzɔ̃dɐʃ] ‘the waves’, *as águas* [ɐzagwɐʃ] and *as almas* [ɐzalmaʃ]. This pattern was transferred into the GGCs.

In his manuscript on Fa d’Ambô, Zamora (MS) brings to attention crucial data showing that in this creole number plays a role not attested in the other GGCs. Some examples are:

- (1) a. *úbêlê d’ôbóio*      udder of a cow (lit. udder of-cow)  
       b. *zúbêlê d’ôbóio*      udders of a cow
- (2) a. *opè bálba*              hair of the beard (lit. foot beard)  
       b. *zópi bálba*             hairs of the beard (lit. foot beard)

Zamora’s phonological explanation for these cases is identical to the explanation above. Since the pluralizing *z* is etymologically related to Portuguese<sup>11</sup>, it raises the question whether a pluralizing rule existed in the proto-GGC or whether this is a local innovation. I will leave this issue for further research.

The general conclusion of this section is therefore that, except for [i]-agglutination in Lung’ie, prosthetic vowels in the GGCs started out as an etymological reflex of the Portuguese article system, which became semantically bleached and generalized as something else than a gender/number system. The gender mismatches that start to show after the oldest diachronic stage support that the Portuguese masculine article became generalized, whereas the use of the agglutinating feminine article (*a*) was abandoned. Less access of these two creoles to Portuguese after diffusion and isolation may have accelerated the loss of the gender-calquing rule.<sup>12</sup>

<sup>10</sup> Note that Cape Verdean Creole exhibits *azágua* [azagwa] ‘rainy season, crops’ (from Ptg *as águas*).

<sup>11</sup> An etymological link to the Spanish definite article system (*lo(s)*, *la(s)*) is unlikely, since we would expect preservation of the initial *l*.

<sup>12</sup> Although the historiography of the island of Annobón is still incomplete in many respects, old documents suggest that the island was settled by slaves from São Tomé, with very limited European presence (e.g. Caldeira 2005).

In the next section I will argue that, in addition to the morphological relation with Portuguese outlined in this section, agglutination in the GGCs is also phonologically conditioned.

### 5. Harmony processes in the GGCs

The GGCs exhibit seven oral vowels<sup>13</sup>, as illustrated in Table 5.

Table 5. Oral vowel inventory of the GGCs.

	<i>Front</i>		<i>Central</i>		<i>Back</i>		
	<b>High</b>	<b>Close-Mid</b>	<b>Open-Mid</b>	<b>Low</b>	<b>High</b>	<b>Close-Mid</b>	<b>Open-Mid</b>
<i>Oral</i>	i	e	ɛ	a	u	o	ɔ

Assimilation in the vowel and consonant system of Santome was first reported by Ferraz (1979), who claims that “[v]owel harmony is a consistent feature of ST phonology.”. Some of Ferraz’ examples of Santome are:

(i) *Vowel Harmony spreading from a tonic vowel* (Ferraz 1979):

	Portuguese	Santome	
(3) Progressive:	pórtɐ	→ pótɔ	‘door’
	fógu	→ fógɔ	‘fire’
(4) Regressive:	rəlózyu	→ ləlózu	‘watch’
	dəgrédu	→ dlegédu	‘exile’

(ii) *Vowel Harmony in clitic groups* (Hagemeijer 2005a: 87)

(5) a.	gɔɔ e (search+3SG)	→ [gɔɔɛ]	‘search it’
b.	volɔ e (get angry + 3SG)	→ [volɔɛ]	‘get angry at him/her’

Note that the type of VH in (5) is restricted to 3sg object pronouns.

Ferraz (1979: 25) describes vowel harmony (VH) in Santome as “a tendency for the same vowel to occur in two consecutive syllables within a morpheme” and provides examples

<sup>13</sup> For the status of nasal vowels, I refer the reader to Günther (1973: 36-7), Ferraz (1979: 20) and Maurer (1995: 23).

for each of the vowels in Table 5 above. In fact, this tendency exists and can be seen in, for instance, cases of epenthesis (e.g. Ptg. *largo* ‘wide’ > Santome *lalugu* ‘wide’). However, these cases should rather be described as instances of vowel assimilation. VH in the GGC is a more restricted phenomenon that occurs in the mid-vowel domain, as follows from Table 6.

Table 6. Dissyllabic words in Santome.

$\begin{matrix} V_2 \\ \backslash \\ V_1 \end{matrix}$	i	u	e	ɛ	o	ɔ	a
i	ligi	migu	izê	mile	libô	jinklo	mina
u	buli	mulu	ubwê	kume	-	-	uswa
e	sêji	dêsu	vêndê	-	-	-	zêma
ɛ	pele	petu	-	vede	-	tebo	bega
o	sôtxi	wôdu	ômê	-	pôvô	-	lôpa
ɔ	doxi	mosu	-	love	-	kodo	bola
a	mali	matu	padê	manse	kasô	avo	faka

Although VH is still an understudied area in the phonology of the GGCs, Table 6 shows that the low and high vowels in general do not impose any co-occurrence restrictions on neighbouring vowels.<sup>14</sup> This, however, stands in sharp contrast with the co-occurrence restrictions on mid-vowels. Very clearly, Santome exhibits a solid rule of mid-vowel stem harmony in at least disyllabic words, meaning that open-mid vowels and close-mid vowels never meet in these cases.<sup>15</sup> Upon inspection of the available literature, I found this rule to apply to Ngola and Lung’ie as well. It is therefore expected to apply to Fa d’Ambu as well, although this distinction cannot be determined from the orthographies used in the available literature (e.g. Barrena 1957, Post 1995).

The tendency toward assimilation, especially in the vocalic domain, raises the question as to whether agglutinated vowels are also subject to constraints imposed on the vowel quality and/or other phonological constraints. It can readily be shown that this is indeed the case. The first observation is that, in agreement with Table 6, mid-vowel stem harmony is always obeyed in the GGCs. If we follow the hypothesis of diachronic layers with respect to agglutination, the following can be observed (x in the table means attested and is followed by an example).

<sup>14</sup> I did not find any data for a high round vowel in the first syllable and a high-mid or low-mid round vowel in the second syllable, which may indicate a rule of vowel elevation.

<sup>15</sup> This situation is also attested in Saramaccan (Norval Smith, p.c.).

Table 7. Agglutination in the oldest diachronic layer (according to the data in Table 1).

V V <sub>aggl.</sub>	a	ɛ	e	ɔ	o	i	u
a	√ (apa)	√ (afe)	-	-	-	-	-
ɔ	√ (omali)	√ (ope)	-	√ (ono)	-	-	-
o	-	-	√ (ôsé)	-	√ (ôlhô)	-	-
u	-	-	-	-	-	-	√ (unu)

Despite the low number of tokens, it follows that in this proto-GGC stage agglutination is strongly associated to vowel height. Low and low-mid vowels in the stem trigger agglutination of a low or low-mid vowel, high-mid vowels occur with high-mid vowels and high vowels occur with high vowels. It is therefore suggestive that Portuguese gender and a reduced ATR-system determined the outcome of agglutinated vowels in proto-GGC. According to Van der Hulst & Van de Weijer (1995: 512), ATR-systems often lack the low [+ATR] vowel or the high [-ATR] vowels, or both.

In the following diachronic stage (Table 2), where the cases of agglutination in Lung'ie and Fa d'Ambu overlap, separate treatment of the data is required.

Table 8. Agglutination in Fa d'Ambu (according to the data in Table 2).

V V <sub>aggl.</sub>	a	ɛ	e	ɔ	o	i	u
a	√ (ampan)	-	-	-	-	-	-
ɔ	√ (olatu)	√ (olemu)	-	-	-	-	-
o	√ (ôman)	√ (ôtesa)	-	-	√ (ôbôyô)	-	√ (ônũa)
u	-	-	-	-	-	-	√ (udum)

Except for *ôtesa* 'forehead', the Fa d'Ambu data in Table 2 are consistent with the patterns found in the oldest diachronic layer (Table 1 and 7). For the etymologically identical items in Lung'ie, however, the situation is quite different, as shown in Table 9.

Table 9. Agglutination in Lung'ie (according to the data in Table 2).

V V <sub>aggl.</sub>	a	ɛ	e	ɔ	o	i	u
o	-	-	√ (ôzê)	-	-	-	-
u	√ (upa)	x (uremu)	√ (ubwê)	-	-	-	-

These data are no longer in agreement with the findings of Table 7 and 8 and strongly suggest that while the cases of agglutination in Fa d’Ambu can arguably still be related to harmony processes found in proto-GGC, [u]-agglutination in Lung’ie was a local development that may have replaced other, previously existing, types of agglutination.

Now consider the agglutination patterns found in items exclusive to Lung’ie, which only involve [u] and [i]-agglutination.

Table 9. Agglutination in Lung’ie (remaining items, cf. Appendix 1).

V V <sub>aggl.</sub>	a	ɛ	e	ɔ	o	i	u
i	-	√ (ite)	√ (izêtxi)	-	-	√ (irixi)	-
u	√ (upanu)	√ (uremu)	√ (upêtu)	√ (usolu)	√ (ufôgô)	-	√ (uzuntu)

Quite clearly, the instances of [i]-agglutination exclusive to Lung’ie clearly involve front harmony with a tonic vowel (i, e, ɛ). However, if this rule is correct, cases such as *upêtu* ‘chest’, *ufew* ‘iron’ or *uwê* ‘eye’ should also receive [i]-agglutination. Therefore, another constraint on the distribution of [i] and [u] is required. This constraint, I argue, is roundness. If there is a front vowel in the stem and no round material (vowels or glides) elsewhere in the stem, [i]-agglutination is triggered; in all the other cases, [u]-agglutination occurs by default.

Cases where Lung’ie exhibits variation, namely *idêntu~udêntu* ‘inside’, *ifi~ufu~ifu* ‘thread, wire’, *usolu~isolu* ‘sun’, *ônôtxi~unôtxi*, *ôzên~uzên* ‘knee’ show that other agglutination patterns in this creole may have been more widespread and became replaced by [u]. In fact, this hypothesis can also be adopted for the differences between Fa d’Ambu and Lung’ie in Table 2. Thus, [u]-agglutination in Lung’ie must be the most recent diachronic agglutination process.

## 6. What triggers harmony processes in the GGCs?

In this section I will discuss the relevance of the different strata that contributed to the formation of the proto-GGC in the phonological conditioning of the agglutination process.

### 6.1. Portuguese

Standard Portuguese exhibits essentially the same oral vowel inventory as the GGCs. VH is well-attested in the GGCs’ lexifier, European Portuguese (e.g. Mateus & d’Andrade 2000), and even more so in Brazilian Portuguese (e.g. File-Muriel 2004). The following types of assimilation are found in Portuguese:

- (i) assimilation in the verbal domain (all conjugations)
- (ii) assimilation in the nominal domain: especially with certain suffixes in BP (height assimilation)
- (iii) metaphony: a tonic vowel harmonizes with post-tonic vowels

However, the processes found in the GGCs are considerably distinct from the processes above, which should therefore be dismissed as the putative trigger.<sup>16</sup>

## 6.2. Western Bantu

The impact of certain Western Bantu languages on the formation of the GGCs has been addressed by Ferraz (1979). In addition to Kongo lexicon in Santome, Ferraz also argues that a number of grammatical features can be assigned to this stratum. The lexical role of Kimbundu (Mbundu) for Ngola is also uncontroversial (Maurer 1992, Lorenzino 1998). As to Lung'ie, which is crucial to the present discussion, Ferraz (1975) briefly discusses the possible substrata for a number of phonological properties in this creole. For instance, he links palatalization and syllable structure to Bantu; voiced velar stops are assigned to the impact of West-African languages other than Bantu. In the light of historical and linguistic evidence, this latter feature must be original from the Nigerian substrate.<sup>17</sup> The presence of trills in Lung'ie and its absence in Santome and Fa d'Ambu, as well as less widespread prenasalization in Lung'ie as compared with Santome or Ngola, is also suggestive of less impact from Bantoid and the retention of more Nigerian-related phonology. At least in the case of syllable structure, Edoid (e.g. Elugbe 1989, Agheyisi 1990) would also account for the patterns observed in Lung'ie. Although Lung'ie shares with the other GGCs some typical Nigerian syntactic features, such as substantial serialization (Hagemeijer 2000, 2005b for Santome), the presence of Bantu is, for instance, felt in the final negation patterns (Güldemann & Hagemeijer 2006, Hagemeijer 2007).

Kongo languages exhibit five oral vowels with phonemic status and significant VH (e.g. Hyman 1999). Ferraz (1979: 49) states that “Kongo exhibits vowel harmony in the same way as ST as a tendency for the same vowel to occur in two consecutive syllables.”. I have claimed above that there is a rule of VH in the GGCs' mid-vowel domain, but no other consistent rule can be found, only cases of vowel assimilation. As to Kongo languages,

<sup>16</sup> Note that in Indo-Portuguese creoles VH is related to Portuguese (Clements 2004).

<sup>17</sup> For instance, some of the items exhibiting co-articulated stop /gb/ in Lung'ie also exist in the other GGCs, but in the form of /kw/ or /bw/, which is arguably related to the massive imports of Bantu slaves on S. Tomé when the island shifts from a *société d'habitation* to a *société de plantation*.

Bentley (1967: 524) mentions the so-called “euphonic preferences”, where [e] patterns with [o] and [i] with [u]. However, Bentley’s use of the term ‘preferences’ should arguably not be related to a ‘tendency’ (in the sense of Ferraz) but to phonological rules that apply with “ideal regularity and consistency”, as Bentley himself points out. In other words, this author refers here to an instance of vowel height harmony, which is well attested across Bantu in general, as well as Kongo languages in particular (Hyman 1999)<sup>18</sup>, where VH between stem and perfective suffixes occurs. It is also obvious that five vowel systems (*i, u, e, o, a*) lack mid-vowel stem harmony.

Since prosthetic vowels in the GGCs typically harmonize with the stressed vowel, it is important to observe how Bantu noun classifiers behave with respect to the stem. According to Hyman (1999), prefix harmonization is uncommon across the Bantu cluster and only applicable to some noun classes. Languages that do exhibit this type of harmony, though, typically have a seven vowel system. The relevant Western Bantu languages do not pattern among this type.

In sum, while it cannot be denied that certain Bantu-related phonological processes do play a role in the GGCs in general, a number of phonological traits in Lung’ie strongly suggest that the impact of Bantu was more reduced in this creole than in the other GGCs. In addition, there is no obvious reason to believe that the type of VH found in the Western Bantu was crucial to pervasive mid-vowel stem harmony and to the observed agglutination patterns.

### **6.3. Edoid**

The impact of Edoid on the formation of the GGCs has been addressed since Ferraz (1979), for lexicon and grammar. In a nutshell, a strikingly high percentage of African lexicon in Lung’ie has its origin in Edoid (Maurer, *forthc.*) and many syntactic properties in the GGCs can be traced back directly to Edoid (Hagemeijer 1999, Hagemeijer & Parkvall 2001; Hagemeijer 2005b, *forthc.*). Although Edoid bears many typological similarities to, for instance, Yoruboid and Kwa languages, both historical and linguistic evidence do not underscore a major role for these clusters in the formation of the GGCs. In fact, the available linguistic and historical evidence strongly suggests that Edoid had a founder effect on the formation of the proto-GGC, preceding in time the massive imports of Bantu slaves from the Congo and Angola.

Nouns in the Edoid cluster are vowel-initial and exhibit either the same oral vowel system as the GGCs or more complex 9/10 vowel systems. VH is substantial in the verb and

---

<sup>18</sup> According to Hyman (1999), southern Kongo languages are unusual in the sense that they exhibit a five vowel system and symmetric height harmony, i.e. harmony whereby *i* and *u* are lowered after *o* and *e*. In asymmetric height harmony, which is very common throughout the Bantu cluster, *u* is only lowered after *o*, but not after *e*.

noun system of many Edoid languages (Delta, South-Western and North-Western Edoid), according to Elugbe (1989). In 7-vowel systems, such as Èdò, however, VH is much more restricted and, when it occurs, it is typically limited to nouns. Elugbe considers VH a feature of reconstructed proto-Edoid. The relevant feature for VH in Edoid is  $[\pm\text{ATR}]$ , yielding the two following mutually exclusive sets of vowels.

(6)	<b>[Expanded]</b>	<b>[Non-expanded]</b>	
	i            u	ɪ            ʊ	
	e            o	ɛ            ɔ	
	ə	a	

Although the 7-vowel chart of the GGCs could not copy a full-fledged ATR-system, it was able to retain the essence of Edoid’s mid-vowel harmony<sup>19</sup>. ATR is also present in the neighbouring language clusters, namely Defoid (Van der Hulst & Van de Weijer 1995: 515-6), Ijoid (Clements & Rialland, to appear; Williamson 1965) and Igboïd (Emenanjo 1987), which makes it an important areal feature. There is at least evidence that these clusters left some lexical imprints on the GGCs. Mid-vowel ATR stem harmony in the GGCs can thus be safely related to Nigerian language clusters with a special role for Edoid.

Furthermore, a noun class system can be reconstructed for proto-Edoid, although most Edoid language only show vestiges hereof (Elugbe 1983, 1989). Èdó, for instance, only exhibits a vowel prefix alternation for only a few items (e.g. òkpà > ìkpà ‘man-men’ (Dunn 1968), whereas Dẹgẹma, a delta-Edoid language, exhibits a complex noun prefixation system for plural formation that obeys rules of VH (Elugbe 1976; Kari 2007). In the typical contemporary Edoid language, however, the vestiges of prefixes do not indicate semantic classes and are the result of “assimilation, vowel harmony, class shifting and levelling” (Elugbe 1983).<sup>20</sup> All the oral vowels can generally be prefixes (i.e. word-initial). In footnote 6, it could be seen that Portuguese words in Èdó are often prefixed by *e* and *ɛ*, which are the only patterns that are not found in the GGCs. In fact, from the available literature, no direct relation semantic or formal relation can be established between prefixation in Edoid and the agglutination patterns in the GGCs. Basically, VH in these languages must have operated as a trigger for the agglutination patterns in the GGCs. In Èdó, for instance, nouns with an [ɔ]-

<sup>19</sup> Note, however, that in Edoid languages with seven-vowel systems, there are only vestiges of VH. For instance, Oḷoma (North-Central Edoid) underwent a reduction from a nine-vowel system to a seven-vowel system, but retains VH (mid-vowels) with singular/plural prefixation (Elugbe 1989).

<sup>20</sup> Rules of VH also apply to past and plurality suffixes (Aikhionbare 1989).

prefix are also likely to have *ɔ*, *ɛ* or *a* in the stem (Elugbe 1989: 78), which is exactly the pattern that obtained in the oldest diachronic stage of the GGCs (see data in Table 1 and the findings of Table 7) and falls into non-expanded in (6) above.

In sum, despite the fact that there are substantial differences between prefixation in Edoid and agglutination in the GGCs, several findings, in addition to other linguistic and historical evidence, favour an Edoid substratum:

- (i) Edoid nouns are obligatorily vowel-initial;
- (ii) Edoid languages generally exhibit ATR harmony or did so in older stages;
- (iii) Edoid languages typically exhibit vestiges of noun prefixation which is (or was) constrained by rules of VH;
- (iv) agglutination is most pervasive in Lung'ie, which is the GGC that exhibits the strongest Edoid impact.

A particular role for Edoid in agglutination may also provide an explanation for the problem presented in section 4, namely the unexpected fact that Fa d'Ambu and Lung'ie have a number of agglutinated items that do not occur in Santome or Ngola. In fact, while it is quite uncontroversial that Lung'ie underwent the strongest impact from Edoid, the impact of Western Bantu was comparatively more significant on São Tomé, not only because of the sugar production but also because of the island's key position in the trans-Atlantic slave trade. The question is therefore whether the proto-GGC exhibited more agglutination at some diachronic stage, or whether a significant part of the agglutination data should be considered coincidental independent innovations in Fa d'Ambu and Lung'ie.

To answer this question, note first that Fa d'Ambu and Lung'ie share several other features not shared by Santome and Ngola, such as the following:

- i) Portuguese nasal diphthong *-ão* yields *-an* in FA and LU and *-on* in ST and NG;
- ii) Several items in FA and LU exhibit /v/ where ST and NG exhibit /b/;
- iii) Less widespread prenasalization in LU and FA than in ST and NG.
- iv) Syntax of the Noun Phrase:
  - Universal quantifier *tudu* 'all' occurs to the right of the head noun in LU and FA and to its left in ST and NG;

- Numerals in FA occur typically to the right of the head noun, which is also the case for ‘one’ and ‘two’ in LU;<sup>21</sup> in ST and NG numerals typically precede the head noun;

I therefore assume that at least a number of common features in LU and FA must have existed in the proto-GGC. Since it can hardly be coincidence that the agglutinated items found in Fa d’Ambu, but not in Santome/Ngola, also exist in Lung’ie, I argue that at least these forms must have existed in (varieties of) the proto-GGC and that the diachrony of the substrata involved in the formation of the GGCs provides an explanation for what might have happened.

I propose that the early proto-GGC is essentially the result of Edoid slaves acquiring Portuguese. Since this period corresponds by and large to the *société d’habitation*, there was arguably better access to the Target Language (TL) and therefore phonetic calquing upon Portuguese gender was additionally able to satisfy the Edoid constraint that nouns are vowel-initial. The fact that the [u], in Lung’ie, and the [ɔ] and [o], in Fa d’Ambu, became so generalized means that the vowel-initial constraint was still active, but also that reanalysis phonetically based on the Portuguese gender distinction was lost and that the Portuguese masculine definite article was retained as input for agglutination.

In this scenario, Lung’ie was the first to branch off the proto-GGC and become isolated at an early stage of creolization and therefore retained a greater number of Edoid features, including agglutination.<sup>22</sup> The direct import of Niger Delta slaves to Príncipe under the contract to Antonio Carneiro (1514-18), and possibly for some years after that, underlines the importance of this area for Lung’ie (Ladhams 2003). This, of course, would account for the heavier influence from Edoid languages in Lung’ie. The fact that a number of words with [u]-agglutination in Lung’ie also have agglutination in Santome and especially Fa d’Ambu, but with a different initial vowel, strongly suggests that [u]-agglutination in Lung’ie locally absorbed previously existing patterns in this creole. This also follows from the items that show variation between [i] and [u]-agglutination.

Fa d’Ambu became isolated at a later stage.<sup>23</sup> Thus, the time span reaching between the isolation of Lung’ie and Fa d’Ambu may well correspond to a decrease in agglutinated

---

<sup>21</sup> Philippe Maurer [p.c.] informs me that all numerals could follow the noun in more archaic varieties of Lung’ie.

<sup>22</sup> Permanent settlement of São Tomé occurred in 1493. The island of Príncipe was officially being settled from 1500 on, but there is documental that some slaves left for the island before the turn of the century.

<sup>23</sup> Settlement of Annobón arguably started in 1503. The island was claimed to have 9 European inhabitants (moradores) in 1507. However, a 16<sup>th</sup> century letter states that Annobón was uninhabited in 1543. Permanent settlement must have occurred in the second half of the 16<sup>th</sup> century.

items in the proto-GGC. Historically, this time span overlaps to a more significant extent with the expansion of the plantation stage (from approx. 1517 on) and the massive arrival of Bantu slaves. However, it is not obvious that Bantu slaves can be held directly responsible for the breakdown of agglutination in the proto-GGC. Nouns, including (singular) prefixes, are to a great extent consonant-initial in the relevant Bantu languages and it can be hypothesized that Bantu speakers learning the proto-GGC related agglutination to the Bantu definite article *o* and therefore reanalyzed the morpheme boundary. It is, however, more likely that Portuguese was responsible for lexical restructuring. Santome, the direct descendant of the proto-GGC, is nowadays the only GGC that exhibits, for instance, widespread consonant clusters, especially with a liquid in C<sub>2</sub>, which is a non-typical feature of Edoid and Western Bantu, but well-attested in Portuguese. Furthermore, Zamora [p.c] informs me that several agglutinated items in contemporary Fa d'Ambu have counterparts that lack the agglutinated vowel, namely *(o)po*, *(ô)bôyô*, *(o)man* and *(am)pan*. It is not clear whether this is a recent process or part of a larger process that decisively affected more items in the past.

Irrespectively of what explanation one adopts for the fact that Santome only exhibits agglutinated items from the oldest diachronic stage, the absence of this feature in Ngola suggests that this creole still remained in contact with Santome after the spread of the Proto-GGC to Príncipe and Annobón.

## 7. Concluding remarks

Although it was shown that agglutination was far more productive in Lung'ie than in any of the other GGCs, only cases involving prosthetic [i] are exclusive to this creole. The remaining patterns, namely with agglutinated [u], [o], [ɔ] and [a], are also attested in the other GGCs. Since both the islands of Príncipe and Annobón were settled from São Tomé, I assume that agglutination must have existed and spread from the proto-GGC.

At least the following diachronic layers should be distinguished. First, etymologically monosyllabic items in the proto-GGC were restructured through agglutinating an initial vowel corresponding to the Portuguese definite masculine and feminine article, following at least phonetically the Portuguese gender distinction, as well as rules of VH from the substrate. Second, agglutination spread to disyllabic items (Lung'ie and Fa d'Ambô) and the Portuguese gender distinction was abandoned, while the rules of VH in the oldest diachronic stage were still active in Fa d'Ambô. Third, Lung'ie exhibits some [i]-agglutination and generalized [u]-agglutination, with an important role for the features [Front] and [Round]. I further argued that in addition to the Portuguese article system, the earliest substrate cluster, Nigerian

language clusters and Edoid in particular played a major role in this process of morpho-phonological restructuring.

### References:

- Agheyisi, R. N. (1990). *A grammar of Edo*. Unesco.
- Aikhionbare, M. O. (1989). "Defining the domain of nasality in Edo". *Studies in African Linguistics*, 20(3), pp. 301-15.
- Baker, Ph. (1984). "Agglutinated nominals in Creole French: their evolutionary significance". *Te Reo* 27: 89-129.
- Barrena, R. (1957). *Gramatica Annobonesa*. Madrid: Consejo de Investigaciones Cientificas.
- Baxter, A. (2004). «The development of variable NP plural agreement in a restructured African variety of Portuguese» in G. Escure & A. Schwegler (eds.) *Creoles, Contact and Language change: Linguistics and social implications*. Amsterdam: John Benjamins, 97-126.
- Bentley, W. Holman. 1967 [1887] *Dictionary and grammar of the Kongo language*. London: Baptist Missionary Society.
- Caldeira, A. (2005). "A 'república negra' de Ano Bom: invenção de um 'Estado' entre duas colonizações". In *Trabalho forçado africano: experiência coloniais comparadas*. Coleção Estudos Africanos. Porto: Campo das Letras - Editores, S.A.
- Clements, C. (2004) "La armonización vocálica en los criollos indo-portugueses", in *Los Criollos de Base Ibérica: ACBLPE 2003* (Eds. M. Fernández, M. Fernández-Ferreiro, N. Vázquez Veiga). Madrid: Vervuert/Iberoamericana, pp. 33-40.
- Clements G. N. & Rialland, A. (to appear). *Africa as a phonological area*. In B. Heine (ed.) *Africa as a linguistic area*. Cambridge: Cambridge University Press.
- Dunn, E. F. (1968): *An introduction to Bini*. East Lansing: African Studies Center, Michigan State University.
- Elugbe, B. (1976). "Noun class vestiges in Dẹgẹma". *Afrika und Übersee* LIX(3). 224-33.
- \_\_\_\_\_ (1983). "Noun class prefixes in proto-Edoid". In Dihoff, I. R. (ed.) *Current approaches to African linguistics*, Vol. 1. Dordrecht: Foris, pp. 59-83.
- \_\_\_\_\_ (1989). *Comparative Edoid: phonology and lexicon*. Delta Series, 6. Port Harcourt: University of Port Harcourt Press.
- Emenanjo, E. (1987). *Elements of modern Igbo grammar: a descriptive approach*. Ibadan: University Press Limited.
- Ferraz, L. Ivens (1975), «African Influences on Principense Creole», Separata de *Miscelânea Luso-Africana*, 153-64.
- \_\_\_\_\_ 1984. The substrate of Annobonese. *African Studies* 43.119-136.
- \_\_\_\_\_ (1979). *The Creole of São Tomé*. Johannesburg: Witwatersrand University Press.

- File-Muriel, R.J. (2004). "An OT approach to vowel height harmony in Brazilian Portuguese". In: Indiana University Linguistic Club Working papers. <https://www.indiana.edu/~iulcwp/pdfs/04-file.pdf>.
- Graham, S & T. Graham (2004). "West Africa Lusolexed Creoles Word List File Documentation". SIL Electronic Survey Reports: <http://www.sil.org/silesr/2004/silesr2004-012.html>
- Granda, G. de (1986). Retenciones africanas en la fonética de criollo portugués de Annobón». *Revista de Filología Románica*, 4, 111-124.
- Grant, A. P. (1995). "Article agglutination in Creole French: a wider perspective". In Baker (ed.) *From contact to creole and beyond*. London: University of Westminster Press, 149-76.
- Güldemann, T. & T. Hagemeyer (2006). "Negation in the Gulf of Guinea creoles: typological and historical perspectives". In Talk delivered at the ACBLPE meeting. Coimbra.
- Günther, W. (1973). *Das Portugiesische Kreolisch der ilha do Príncipe*. Marburg an der Lahn.
- Hagemeyer, T. (1999), «As Ilhas de Babel: a criouliização no Golfo da Guiné». *Revista Camões*, 6, 74-88.
- \_\_\_\_\_ (2005a). "Going in the clause: *ba* and *be* in Santome", *Journal of Portuguese Linguistics*, 3-2. Lisboa: Colibri.
- \_\_\_\_\_ (2005b). "The origins of serialization in the Gulf of Guinea creoles", talk delivered at Creole language structure: between substrates and superstrates, Leipzig.
- \_\_\_\_\_ (2007). *Clause structure in Santome*. Doctoral Dissertation, Universidade de Lisboa.
- Hagemeyer, T. & M. Parkvall (2001). The development of the Gulf of Guinea creoles. Talk delivered at the ACBLPE annual meeting, Coimbra.
- Holm, J. (1988). *Pidgins and creoles*, Vol. 1 (*theory and structure*). Cambridge: Cambridge University Press.
- Hulst, H. van der & J. van de Weijer (1995a). Vowel harmony. In J.A. Goldsmith (ed.) *The handbook of phonological theory*. Cambridge, MA.: Blackwell, 495-534.
- Hyman, L.M. (1999). The historical interpretation of vowel harmony in Bantu". In J.M. Hombert & L.M. Hyman (eds), *Bantu historical linguistics: theoretical and empirical perspectives*. Stanford: CSLI, 235-295.
- Kari, E. (2007). "Vowel harmony in Degema, Nigeria". *African Study Monographs*, 28(2), 87-97.
- Ladhams, J. (2003). *The Formation of the Portuguese Plantation Creoles*. PhD dissertation, University of Westminster.
- \_\_\_\_\_ (forthc.). "Article Agglutination and the African Contribution to the Portuguese-based Creoles". In Bartens, A. & Ph. Baker (eds.) *Black through White*. London: Battlebridge Publications.
- Lorenzino, G. (1998). *The Angolar Creole Portuguese of São Tomé: its Grammar and Sociolinguistic History*. PhD Dissertation, City University of New York.
- Mateus, H. & d'Andrade, E. (2000). *The phonology of Portuguese*. London: Oxford University Press.

- Maurer, Philippe (1992). "L'apport lexical bantou en Angolar", em: *Afrikanische Arbeitspapiere* 29, 163-174.
- \_\_\_\_\_. (1995). *L'Angolar: un Créole Afro-Portugaise parlé à São Tomé*. Hamburg: Helmut Buske Verlag.
- \_\_\_\_\_. (forthc.). Lung'ie. The Afro-Portuguese Creole of Príncipe Island (Gulf of Guinea). Grammar – Texts – Vocabulary. London: Battlebridge.
- Melzian, Hans Joachim (1937). *A concise dictionary of the Bini language of southern Nigeria*. London: Kegan Paul, Trench Trubner & Co.
- Negreiros, A. (1895). *Historia Ethnographica da ilha de S. Tomé*. Lisboa: 1895.
- Parkvall, M. (2000). *Out of Africa*. London: Battlebridge Publications.
- Post, M. (1995). "Fa d'Ambu", In Arends et alii (eds.) *Pidgins and creoles: an introduction*. Amsterdam/Philadelphia: John Benjamins, pp. 191-204.
- Rougé, J.-L. (1992). «Les langues des Tonga», in *Actas do colóquio sobre crioulos de base lexical portuguesa*. Lisboa: Colibri.
- Schang, E. (2000). *L'émergence des Créoles Portugais du Golfe de Guinée*. Docotoral Dissertation, Université de Nancy 2.
- Schwegler, A. (2006). "Bantu elements in Palenque (Colombia): Anthropological, archeological and linguistic evidence", em Havisser, Jay B. (ed.): *Confronting social issues of the African diaspora*. London: Routledge Press.
- Williamson, K. (1965). *A grammar of the Kolokuma dialect of Ijọ*. West African Language Monographs 2. Cambridge: Cambridge University Press.
- Zamora, A. (MS). *Gramática descriptiva del fa d'ambô*.

Appendix 1. Prosthetic vowels in the GGC.<sup>24</sup>

<i>Lung'ie</i>	<i>Santome</i>	<i>Fa d'Ambu</i> <sup>25</sup>	<i>Lunga Ngola</i>	<i>Portuguese etymology</i>	<i>Meaning</i>
afe	afe	fe		<i>fê</i>	'faith'
apa	apa	[pala]	<b>anpa</b>	<i>pá</i>	'spade, shovel'
(gêêza)-jase	ase	[geza]	[ngeedha]	<i>sé</i>	'church, cathedral'
[nomi]	aglasa	galasa		<i>graça, nome</i>	'name'
ope	ope	ope	<b>ope</b>	<i>pé</i>	'foot, leg'
opo	opo	opo	<b>ompo</b>	<i>pó</i>	'dust'
ose	ose	ôsê	<b>onthe</b>	<i>céu</i>	'sky, heaven'
ono	ono	ono	[numbu]	<i>nó</i>	'knot'
odo	odo	dól	<b>odo~oro</b>	<i>dó</i>	'mourning'
ônôtxi ~ unôtxi	nôtxi	notsyi	[n'thuku]	<i>noite</i>	'night'
ôryô 'river'	ôlhô <sup>26</sup>	[lubela]	awa	<i>rio</i>	'river'
ôzê (n)~ôzên~uzên	zê	<b>ôzôyô</b>	[m'puna~puna]	<i>joelho</i>	'knee'
uari	<b>oali</b>	oventu	ventu	<i>ar, vento</i>	'air'
ubaaku	blaku, kobo	[fuladu]	[kobo]	<i>buraco, furado, cova</i>	'hole'
ubaasu	[mon]	bazu, baasu	[mon]	<i>braço, mão</i>	'upper arm'
ubanku	banku	banku	banku	<i>banco</i>	'bank'
ubasu	basu	basu	[kosi]	<i>baixo</i>	'under(neath)'
ubudu	budu	budu, buudu	buru	?	'stone'
ubuka	boka	boxo	boka	<i>boca</i>	'mouth'
udedu	dedu	dedu(-oman)	reru	<i>dedo</i>	'finger, (toe)'
udôdô	dôdô	dôdô	[toto]	<i>doido, tótó</i>	'idiot'
udumu	dumu	<b>udum</b>		<i>Edo (u)dumu</i>	'pestle'
udyabu	[demono]	[damon(o)]	[romono]	<i>diabo, demónio</i>	'devil'
ufaka	faka	faka	faka	<i>faca</i>	'knife'
ufatu	fatu	[bitsyidu]		<i>fato, vestido</i>	'suit'
ufôgô / ufôgu	fôgô	fogo	fôgô	<i>fogo</i>	'fire'
ufundu	fundu	fundu	fundu	<i>funfo</i>	'depth'
ugaafu	galufu	gefa	ngalufu	<i>garfo</i>	'fork'
ugagu	gagu	gagi	ngagu	<i>gago</i>	'stutterer'
ugalu	galu	galu	[n'kombo]	<i>galo</i>	'cock'
ugatu	gatu	gatu, ngatu		<i>gatu</i>	'cat'
uguya	guya	guya, gunha	ngunha	<i>agulha</i>	'needle'
ukabu				<i>cabo</i>	'end'
ukagu	[luge]	xálgu		<i>cargo, lugar</i>	'responsability'
ukalu	kalu	ekuza		? <i>caldo</i>	'broth'
ukampu	kampu		kampu	<i>campo</i>	'field'
ukanu	kanu			<i>cano</i>	'pipe'
ukantu	[kantxin]		[tia]	<i>canto</i>	'corner'
ukaru	karu	[moto]	karu	<i>carro</i>	'car'
ukuru	kulu	kulu	[n'thuku]	<i>escuro</i>	'darkness'
ukuru	kulu	kulù	kuru	<i>cru</i>	'raw'
ukwatu	kwatu	[xotchian]		<i>quarto</i>	'room'
uladu	[bodo]	ladu		<i>lado</i>	'side'
ulasu	lasu	lasu	[n'vumbu]	<i>laço</i>	'knot'
ulalu	lalu	lalu		?	'skin disease'
ulensu	lensu	lensu(-zubela)		<i>lenço</i>	'handkerchief'
umari, umwe(n)	<b>omali</b>	<b>omal</b>	[mionga]	<i>mar</i>	'sea'
umatu	matu	<b>omatu</b> , matu	matu	<i>mato</i>	'bush-bush, jungle'
umundu	mundu	mundu		<i>mundo</i>	'world'
unfenu	nfenu	<b>onfenu</b>	nfenu	<i>inferno</i>	'hell'

1 Sources: Lung'ie - Günther (1973), Maurer (forthc.); Santome – Ferraz (1979), my own corpus; Fa d'Ambu – Barrena (1957), Graham & Graham (2004), De Granda (1986), Marike Post [p.c.], Armando Zamora [p.c.]; Angolar – Maurer (1995). Fricative [ʃ] is represented as *x* in Santome, Lung'ie and Ngola. In Fa d'Ambô, [ʃ] is *ch* and [x] is *x*. The bracketed items are not related to the etymology of the agglutinated form.

<sup>25</sup> I have drawn the Fa d'Ambu from written sources, namely Barrena (1957), De Granda (1986) and Graham & Graham (2004). However, the data pieces are often contradictory, especially with respect to the vowel qualities. I have therefore submitted these forms to a native speaker, Armando Zamora, whose judgements are ultimately reflected in the table.

<sup>26</sup> Form attested in Negreiros (1895)

<i>Lung'ie</i>	<i>Santome</i>	<i>Fa d'Ambu</i> <sup>27</sup>	<i>Lunga Ngola</i>	<i>Portuguese etymology</i>	<i>Meaning</i>
upanu	panu	panu	panu	<i>pano</i>	'cloth'
upêtu	petu	[boxo-kusá]	[pota-kotho]	<i>peito</i>	'chest'
urabu	labu	<b>labu</b>	[inkila~n'kila]	<i>rabo</i>	'tail'
uramu	[aba]	[aba]	[n'thala, tango]	<i>ramo</i>	'branch'
uratu	latu	<b>ôlatu</b>	latu	<i>rato</i>	'rat, mouse'
uremu	lemu	<b>olemu</b>	lemu~lêmu	<i>remo</i>	'oar'
usaku	saku	saku	thaku	<i>saco</i>	'bag'
usalu	salu	salu	thalu	<i>sal</i>	'salt'
usolu (isolu)	solo	solo	tholo	<i>sol</i>	'sun'
usuva	suba	[awa]	thuba	<i>chuva</i>	'rain'
utabu	taba	taba	taba	<i>tábua</i>	'plank'
utasu	tasu	tasu	tathu	<i>tacho</i>	'bowl'
utempu	tempu	tempu	tepu	<i>tempo</i>	'weather'
uventu	ventu	<b>oventu</b>	ventu	<i>vento</i>	'wind'
uzuntu	zuntu	zuntu	[dhunta]	<i>junto</i>	'next to'
ubaw	balu	[lama]	[mavu]	<i>barro</i>	'soil, clay'
ubwê	bwê	<b>ôbôyô</b>	buê	<i>bói</i>	'cow'
ubên	bên	ben		<i>bem, riqueza</i>	'richness, possessions'
ufew	felu	<b>onfelu</b>	felu~fêlu	<i>ferro</i>	'iron'
uga	lwa	<b>ôlua</b>	lua	<i>rua</i>	'street', 'outside'
ufya	fya	fala	fia	<i>folha</i>	'leaf'
uku	<b>uku</b>	[zanga]	[ndagu]	<i>cu</i>	'arse'
umpan	mpon	<b>ampan</b>	pon	<i>pão</i>	'bread'
unu	<b>unu</b>	[dodyí]		<i>nu</i>	'naked'
unwan	nŵa	<b>ônŵa, ônŵya</b>	[mbêi~mbêzi~mêzi]	<i>lua</i>	'moon'
uman	mon	<b>ôman, man</b>	mo~mon	<i>mão</i>	'hand, lower arm'
upa	po	<b>ôpa</b>	po	<i>pau</i>	'stick, tree'
usan	son	san	thon	<i>chão</i>	'ground'
uwê	wê	ôyô	wê	<i>olho</i>	'eye'
idintxi	dêntxi	denchi	rêtxi~dêtxi	<i>dente</i>	'tooth'
ifi (ufi, ufu)	fi	filu	fi	<i>fio</i>	'thread, wire'
ijinjibi	jjimpli	yimbil		<i>gengiva</i>	'gum'
imin	min	milu	minhu	<i>milho</i>	'maize'
irixi	lix	lichi	lisi~risi~disi	<i>nariz</i>	'nose'
ise	sela	[xama]	[kama]	<i>esteira, cama</i>	'bed'
isengi	sangi	sangi	thangi	<i>sangue</i>	'blood'
ite	tela	[lama]	tela	<i>terra, lama</i>	'soil, clay'
ivin	vin	vin	vi	<i>vinho</i>	'wine'
ixima	xima <sup>28</sup>	[liba]	[riba~diba]	<i>(em?) cima, arriba</i>	'on top'
ixize	xinza	chinza	txindja	<i>cinza</i>	'ashes'
izêtxi	zêtê	zete(-palm)	[mazi]	<i>azeite</i>	'palm oil'
idêntu (udêntu)	glêntu	dentulu, dantulu	lêtu	<i>dentro</i>	'inside'
ipin	pyan	pina, pyan	pinha	<i>espinha</i>	'thorn'
tesa	tesa	<b>ôtesa</b>	[kwanda polo]	<i>testa</i>	'forehead'

<sup>27</sup> I have drawn the Fa d'Ambu from written sources, namely Barrena (1957), De Granda (1986) and Graham & Graham (2004). However, the data pieces are often contradictory, especially with respect to the vowel qualities. I have therefore submitted these forms to a native speaker, Armando Zamora, whose judgements are ultimately reflected in the table.

<sup>28</sup> By far the most common form is *liba*, but *xima* occurs a few times in my corpus.