

Vowel Harmony in the Proto-Creole of the Gulf of Guinea

Abstract

The proto-Creole of the Gulf of Guinea is the common ancestor of the Santome, Angolar, Lung'Ie and Fa d'Ambô languages. Based on a supposed occurrence of vowel harmony (vH) in these modern languages and the proven influence of the languages of the Niger Delta (HAGEMEIJER, 2009), in which vH is attested, this study aims to discuss the existence of vH processes with [ATR] mid-vowels in the proto-Creole of the Gulf of Guinea, considering as a starting point its phonological and lexical reconstruction (BANDEIRA, 2017). Therefore, we intend to verify what phonological parameters related to vH in the literature (trigger, target, domain, application direction and blocking element) indicate about vH processes in proto-Creole, based on the lexical reflexes of a set of cognates in daughter languages and proto-forms. We will show that many of the items interpreted as harmonic derive from the maintenance of mid-vowels in Portuguese etyma and from the insertion of a vowel copy in order to avoid roots ending in consonants in the proto-Creole of the Gulf of Guinea. In this analysis, we will argue for the hypothesis according to which parasitic vH processes (COLE; TRIGO, 1988) with [ATR] mid-vowels in contiguous syllables occurred in proto-Creole within the prosodic word domain. However, the vH was limited because, based on the observation of proto-forms and harmonisation patterns with [ATR] mid-vowels in contiguous syllables, there were many cases in which there was neutralisation of final non-stressed mid-vowels and, therefore, no application of the harmony process in favouring contexts.

Keywords: Vowel harmony; Proto-Creole of the Gulf of Guinea; Phonology.

Resumo

O proto-crioulo do Golfo da Guiné (PGG) é o ancestral comum do santome, angolar, lung'Ie e do fa d'Ambô. Considerando a controversa existência de harmonia vocálica (HV) nessas línguas modernas e a influência das línguas de substrato da região do Delta do Níger (HAGEMEIJER, 2009), nas quais há HV, este trabalho pretende discutir a existência de processos

de HV no PGG, tomando como ponto de partida a reconstrução fonológica e lexical proposta por Bandeira (2017). Portanto, nossa intenção é verificar o que a aplicação dos parâmetros relacionados à HV na literatura (gatilho, alvo, domínio, direção de aplicação e elemento de bloqueio) nos permite inferir sobre os processos de HV no PGG, baseado nos reflexos lexicais de um grupo de cognatos nas línguas-filhas e nas proto-formas. Será mostrado que muitos dos itens considerados harmônicos derivam da manutenção de vogais-médias já 'harmônicas' nos étimos portugueses e da inserção de vogais cópias em final de palavras do étimo português, cujo objetivo era evitar raízes terminadas em consoantes nas proto-formas no PGG. Nessa análise, defenderemos que processos de harmonia vocálica parasítica (COLE; TRIGO, 1988) com vogais médias [ATR] em sílabas contíguas ocorriam no proto-crioulo no domínio da palavra prosódica. No entanto, a partir da observação das proto-formas e dos padrões harmônicos com as vogais médias [ATR] em sílabas contíguas, mostraremos que a HV era limitada pois havia muitos casos de não-aplicação do processo, principalmente quando a neutralização de vogais médias finais átonas descaracterizava o alvo para o processo.

Palavras-chave: Harmonia vocálica; Proto-crioulo do Golfo da Guiné; Fonologia.

1. Introduction

The proto-Creole of the Gulf of Guinea, which emerged on São Tomé Island in the late 15th and early 16th centuries, at the beginning of the Portuguese colonisation of that region, is the common ancestor of the Santome (ISO code 639-3: CRI), Angolar (AOA), Lung'Ie (PRE) and Fa d'Ambô (FAB) languages spoken in the region of the Gulf of Guinea, West Africa. The isolation and the displacement of some populations of São Tomé Island, as well as new linguistic contributions from African languages (deriving from the constant renewal of the incoming enslaved population), promoted the speciation of proto-Creole. On the one hand, Santome developed in colonisation nuclei in São Tomé Island, and Angolar is the language of maroon communities that arose from descendants of runaway slaves who escaped from mills, villages and plantations. On the other hand, proto-Creole speakers were taken

to the Príncipe Island and Annobón Island, where the conditions for the speciation of Lung'Ie and Fa d'Ambô were met, respectively (BANDEIRA, ARAUJO, FINBOW, 2019). Currently, Santome, Angolar and Lung'Ie are spoken in the Democratic Republic of São Tomé and Príncipe, and Fa d'Ambô is spoken on the Annobón and Bioko Islands, territories of the Republic of Equatorial Guinea.

Bandeira (2017) proposed a phonological and lexical reconstruction of the proto-Creole of the Gulf of Guinea (PGG) based on a set of 536 cognates of CRI, AOA, PRE and FAB (henceforth referred to as the daughter languages). The data from the daughter languages were collected during original fieldwork and from the literature (MAURER, 1995, 2009; LORENZINO, 1998; SEGORBE ZAMORA, 2010; ARAUJO, HAGEMEIJER, 2013). According to Bandeira (2017, p. 274) and Bandeira, Araujo and Finbow (2019, p. 3), the consonantal system of PGG had 18 consonants¹: *p, *b, *t, *d, *k, *g, *gb, *f, *v, *s, *z, *m, *n, *ɲ, *l, *r, *w and *j. The vowel system of PGG, in turn, consisted of seven oral vowels: *i, *e, *ɛ, *a, *ɔ, *o and *u. Additionally, there were no nasal vowels; instead, there was a spreading process from nasal consonants in coda conditioning vowel nasalisation (BALDUINO et al., 2015; AGOSTINHO, 2016). Besides, V and CV syllables, PGG allowed complex onsets (*C-l, *C-w or *C-j). Coda position could be filled by liquid consonants (*r, *l), approximant consonants (*w, *j), the fricative (*S) and an unspecified nasal consonant (*N). Stress was usually placed on the penultimate syllable in nominal words, shifting to the last when the syllable was heavy. However, there were also occurrences of final stressed words ending in light syllables and

¹ Notation: An asterisk (*) precedes all proto-phonemes and proto-forms. Daughter language items can be presented in two ways: (i) according to their phonetic transcription, and (ii) according to their phonological representation. The first representation is indicated by brackets [], as in ['kaku] *caco* 'fragment', in PRI, the latter is indicated in /bwe/ 'ox' in CRI. The primary stress was marked by the symbol (') preceding the tonic syllable as in ['budu] 'stone' in AOA. The glosses are indicated after the given word in italics (in Portuguese) and between single quotation marks (example: 'stone') (in English). The symbols used for transcription are in accordance with the International Phonetic Alphabet.

antepenultimate stressed words.

In this paper, our goal is to discuss the existence of vowel harmony (vH) in [aATR] mid-vowels in the proto-Creole of the Gulf of Guinea based on its phonological and lexical reconstruction (BANDEIRA, 2017). We will argue for the following hypothesis: Parasitic vH processes (COLE; TRIGO, 1988) in [aATR] mid-vowels in contiguous syllables occurred in a limited way in proto-Creole within the domain of prosodic words. Thus, based on the lexical reflexes of a set of cognates of the daughter languages and proto-forms of PGG, we intend to verify the usual phonological parameters related to vH (trigger, target, domain, application direction and blocking element).

The text is organised as follows: In section 2, we define the term *vowel harmony* and discuss how vH has been treated in the literature of the daughter languages. In 2.1, we show that in PGG, vowel harmony by copy must be separated from classical [ATR] mid-vowel harmony. Then, based on proto-forms of Portuguese etyma with [+ATR] vowels, section 2.2 highlights that not every case of [aATR] mid-vowels in contiguous syllables should be considered a result of an application of a vH process. In section 2.3, [aATR] mid-vowel harmony in PGG will be addressed, as well as its non-application despite favourable contexts. In Section 2.4, unexpected non-harmonic data is discussed, whereas section 2.5 focus on compounds in PGG. Finally, in section 2.6, non-Portuguese etyma are addressed. Section 3 presents the final remarks.

2. Vowel harmony

Vowel harmony is a phonological process that can be defined as the phonetic influence of one vowel over another within a given domain (PULLEYBLANK, 1988; MOHANAN, 2009; NEVINS, 2010). As a phonological process, harmony requires a triggering element, a target, a domain (that is, an environment in which the rule applies) and an application direction, which can be to the right (progressive harmony),

to the left (regressive harmony) or in both directions. The harmony process is visible when there is a recurrent agreement of vowel features of the same quality in root/word vowels or when there is a change in the nature of a vowel element influenced by a trigger vowel in a root or an affix that affects one or more vowels in the root or affixes. In (1a), there are some examples of [aATR] mid-vowel harmony in Angolar, a language spoken in São Tomé. Affix-root harmony (1b) can be observed in Akan, a language spoken in Nigeria. In this language, the third-person singular prefix vowel is the target of a regressive mid-vowel harmony process triggered by an [ATR] feature from the root. Therefore, in (1) the prefix vowel is [o] or [ɔ], depending on the [ATR] feature of the root vowel.

(1) (a) [aATR] mid-vowel harmony in Angolar (ARAUJO, BANDEIRA, 2019)

'vɛɛ 'green'	'ɔsɔ 'bone'	ɔ'pɛ 'foot'
θe'ke 'sand'	'fogo 'fire'	'sore 'soldier'

(b) Prefixal harmony contrast in Akan (OLA, 2001; O'KEEFE, 2003; NEVINS, 2010):

a. ɔ-bɛku	'3SG-fight'
b. o-beku	'3SG-dig'

The existence of vowel harmony in Portuguese-based Creole languages of the Gulf of Guinea has been advocated by Ferraz (1979) for Santome, by Maurer (1995) for Angolar and by Segorbe Zamora (2010) for Fa d'Ambô. Hagemeyer (2009, p. 36–37), on the other hand, suggests that a type of harmony of [ATR] mid-vowels can be found in all four languages (See Table 1). However, Agostinho (2016, p. 130) states that there is no vowel harmony in modern Lung'Ie. Table 1 shows the specified features of PGG vowels, from which we will address the vowel harmony of mid-vowels.

	i	e	ɛ	a	ɔ	o	u
[high]	+	-	-	-	-	-	+
[low]	-	-	-	+	-	-	-
[back]	-	-	-	+	+	+	+
[ATR]	+	+	-	-	-	+	+

Table 1: Vowel chart of PGG.

Based on the available literature, a comprehensive characterisation of the vowel harmony processes in Santome, Angolar, Lung'Ie and Fa d'Ambô has been problematic. Ferraz (1979, p. 49) states, in relation to Santome, that vowel harmony is the 'tendency of the same vowel to occur in two consecutive syllables within a morpheme'. With regard to Angolar, Maurer (1995, p.36) argues that when two vowels are contiguous in a sentence, a progressive assimilation phenomenon may occur. Segorbe Zamora (2010, p. 63), in turn, defines vowel assimilation as a process in which the vowel segments of a word may present a change of timbre, becoming another vowel. Vowel assimilation in Fa d'Ambô, according to Segorbe Zamora, can be very frequent within either a word or a sentence. However, the author does not specify a prosodic constituent as a domain. Hagemeyer (2009, p.37) states that '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 co-occur in these cases'. Additionally, Hagemeyer (2009, p. 37) points out that 'Mid-vowel ATR stem harmony in the G[ulf of] G[uinea] C[reole]s can (...) be safely related to Nigerian language clusters with a special role for Edoid'.

To propose vowel harmony processes for the four daughter languages—that is, vowel changes at a word level—the literature considers, on the one hand, the dynamics of vowel combinations within a same word or compares Portuguese etyma with synchronic forms of Santome, Angolar, Lung'Ie and Fa d'Ambô. On the other hand, it

considers the supposed absence of pairs with [aATR] disharmonic vowels in contiguous syllables within each language. We will follow Viaro (2011, p. 99), who defines the term 'etymon' as 'the equivalent form of a same word immediately preceding in any past synchrony'. However, this study is based on reconstructed proto-forms, a phonological reconstruction of proto-Creole and on the contemporary data for CRI, PRE, AOA and FAB, and it aims to discuss the nature of vowel harmony processes in proto-Creole. Thus, the focus of this study shifts from vowel harmony to the linguistic ancestor of the immediately preceding cluster (i.e., proto-Creole); therefore, mistakes in the analysis of lexical items via a Portuguese etymon alone are avoided. To state that a certain PGG word has a Portuguese etymon means arguing that this PGG word may be related to Portuguese, albeit without focusing on the origin of that word itself. At the same time, we intend to verify which phonological parameters related to vowel harmony (trigger, target, domain, application direction and any blocking element) can be found in PGG and in the lexical reflexes of a set of cognates of daughter languages.

From a *corpus* of 536 PGG proto-forms (BANDEIRA, 2017, p. 370–416), we selected all 395 nonverbal lexical items. Among them, 66 proto-forms (16.7% of the total) present sequences of [aATR] mid-vowels in contiguous syllables. However, these sequences do not necessarily constitute examples of the application of [aATR] vowel harmony. For this reason, the 66 proto-forms were classified into three groups: (i) 14 proto-forms with a vowel copy (21.2% of the total), (ii) 23 proto-forms with [+ATR] mid-vowels in contiguous syllables with Portuguese etyma (34.8%) and (iii) 23 proto-forms with [aATR] mid-vowels in contiguous syllables not related to groups (i) and (ii). In addition, five proto-forms did not have a Portuguese etymon. [ATR] vowel harmony differs from that of vowel copy but does not exclude it, as will be shown in section 2.1. Its occurrence is linked to the dissolution of forbidden syllabic types, as well as the mere coincidental

presence of [a_{ATR}] vowels in contiguous syllables already present in Portuguese or non-Portuguese etymon forms. Next, each of the three groups will be addressed separately.

2.1 Vowel copy

Based on Bandeira's proto-form data, we observed that vowel copy is a phonological process in proto-Creole. The proto-forms in (2)–(5) are representative of vowels in contiguous syllables generated by vowel copy, whose function seems to have been to prevent words with two or more syllables² from ending in the consonants /r/, /l/ or /S/. Nevertheless, vowel copy differs from vowel harmony. Any vowel can be copied; thus, vowel quality is not solely associated with [ATR] features, which must display agreement, as can be seen in (2), in which there is a copy of the vowels [u] and [i] from the vowel of the stressed syllable in the lexical items *'zulu and *du'mini, respectively. In addition, vowel insertion is linked to the tendency to avoid words ending in consonants in PGG. In the copy process, a vowel identical to the trigger vowel, in a stressed syllable (i.e., in a strong position (BECKMAN, 1998)), was inserted to the right of the final consonant in the word, allowing the formation of a new syllable³ ending in a vowel in harmony with the vowel existing in the Portuguese etymon.

- (2) a. *'zu.lu < azul 'blue'
 b. *du.'mi.ni < dormir 'to sleep'
 c. *'ri.si < nariz 'nose'

As for mid-vowels, 10 lexical items of our corpus contain a copy of [o], three of [ɔ] and one of [ɛ], as shown in (3)–(5). All vowel copies

² Among the 395 nominal lexical items of the *corpus* collected by Bandeira (2017), 10 (2.53%) are monosyllables; however, all end in vowels—such as *'lwa 'street' and *we 'eye'—except for a single case, the word *'seiS 'six', which ends in a consonant.

³The division of syllables is represented by a dot.

were inserted in the same context: the right edge of a final stressed syllable from an etymon ended in a consonant. The insertion allowed the creation of a new syllable, in which the consonant of the coda was resyllabified, becoming part of the onset, and formed a new syllable with the vowel copy.

- (3) a. *'ko.lo < *cor* 'colour'
- b. *'do.lo < *dor* 'pain'
- c. *ka.'lo.lo < *calor* 'heat'
- d. *fa.'vo.lo < *favor* 'favour'
- e. *do.'to.lo < *doutor* 'physician'
- f. *a.'ro.so < *arroz* 'rice'
- (4) a. *'sɔ.lɔ < *sol* 'sun'
- b. *'pjɔ.rɔ < *pior* 'worse'
- c. *'zɔ.lɔ < *anzol* 'hook'
- (5) a. *'mɛ.lɛ < *mel* 'honey'

Given the above, we may state that the process of vowel copy occurred to avoid roots ending in consonants. Thus, vowel copy can be considered a broad subtype of vowel harmony in proto-Creole because all vowels could be entirely copied. However, vowel copy differs from the harmony process that occurs with [ATR] mid-vowels, as will be demonstrated.

2.2 Proto-forms with [+ATR] mid-vowels in the Portuguese etyma

The 23 proto-forms with [+ATR] mid-vowels in contiguous syllables identical to those in the Portuguese etyma correspond to almost 35% of all 66 harmonic lexical items in the *corpus*. Due to socio-linguistic pressures at its genesis, PGG was in contact with Portuguese, which, in turn, is a language that has lexical items with identical [+ATR] mid-vowels in contiguous syllables. However, this is not linked to a process

of vowel harmony in this language. Hence, to conduct a comprehensive analysis of vowel harmony in PGG, it is important to eliminate lexical items with a Portuguese etymon with identical [+ATR] mid-vowels in contiguous syllables based on the list of examples of VH, because such words already had such 'harmonic' vowels in the lexifier language. Of the 23 proto-forms, there are occurrences of vowels with the same quality in relation to the [+ATR] feature. However, there are no examples of non-agreement regarding this feature because the lexical items with identical contiguous vowels with a Portuguese etymon in the *corpus* contained only [+ATR] mid-vowels, which were kept in the PGG.

- (6) a. *'zete < *azeite* 'oil'
 b. *feve'relu < *fevereiro* 'February'
 c. *'oro < *ouro* 'gold'

The data in (6) do not suggest a change of [ATR] features⁴ in the adaptation from the Portuguese etymon to the proto-form. If there were any modifications, it would be a strong indication of a vowel harmony process, in which the features of a vowel underwent changes, and the proto-form would be listed as a case of VH and not as a case of faithfulness to the etymon.

The PGG proto-forms presented in sections 2.1 and 2.2 showed that not all examples with identical mid-vowels are cases of [ATR] vowel harmony; there were cases of words with harmonic [+ATR] vowels in the Portuguese etymon incorporated without a vowel change into the PGG, and vowel copy processes triggered by an element to avoid words ended in consonants. In the next section, we will discuss cases of proto-forms with harmonic vowels whose [+ATR] quality cannot be attributed to a Portuguese etymon or to a broad vowel copy process.

⁴ In the data, other processes also occur, but they are irrelevant to the present discussion.

2.3 Vowel harmony with [aATR] mid-vowels in the PGG

Our *corpus* has 23 nominal lexical items that display [aATR] mid-vowel agreement and five proto-forms with a non-Portuguese etymon reconstructed with [aATR] vowels. Thus, these 23 or 28 nominal proto-forms present a compulsory agreement of the [ATR] feature in the mid-vowels of contiguous syllables (i.e., instances of real *vh*). Before we address in detail these 23 or 28 proto-forms, we present Table 3, in which all possible vowel combinations in contiguous syllables in PGG are shown. In Table 3, a dash (—) is used to show impossibility, whereas a question mark (?) is used to show that data was not found in the *corpus*, although the combination is theoretically possible. We conclude that PGG allowed contiguous syllables with combinations of various types of vowels, except items with non-agreeing [ATR] mid-vowels in contiguous syllables. Therefore, one can eliminate other types of harmony features, such as roundness and height. Thus, questions on [ATR] mid-vowel harmony in daughter languages (HAGEMEIJER, 2009) are also relevant to PGG.

V1/V2	*i	*u	*e	*ɛ	*o	*ɔ	*a
*i	*'risi	*'tligu	*i'ze	*'ine	*'tio	?	*'tliɸa
*u	*'sukli	*'pulsu	*su'pe	*ku'mɛ	*o'zoɸu	?	*'guɸa
*e	*'deNti	*'teru	*'lete	—	*a'mole	—	*'pena
*ɛ	*'pɛli	*'pletu	—	*'mɛɛ	—	*vɛ'gɔɸa	*'bega
*o	*'koni	*'wotu	*klo'ze	—	*'mojo	—	*'sopa
*ɔ	*'mɔli	*gɔr'dura	—	*ɔ'pɛ	—	*'kɔblo	*kɔ'razi
*a	*ba'li	*'salu	*ba'te	*ma'rɛlu	*ka'so	*aN'tɔNtɛ	*'bana

Table 3: Combinations of vowels in words with two or more syllables in PGG.

Of the 23 proto-forms in which there is agreement of [ATR] features and the etyma of which are Portuguese, 21 contain a [-ATR] mid-vowel and therefore must be treated as innovations of PGG, because Portuguese did not present lexical items in the 15th and 16th centuries with two or more [-ATR] mid-vowels in contiguous syllables.

In (7), there are three sets of data in which the result was a combination of [-ATR] mid-vowels (such as [ɔ...ɛ]) in contiguous syllables from etyma in which there was a [-ATR] vowel in a tonic position, as in (7-a) and (7-b), which triggered a valence change in the [+ATR] feature in the vowels of contiguous syllables. Thus, the *vh* process applied categorically when there was an [-ATR] mid-vowel on the stressed syllable in the Portuguese etymon. However, according to the data in (7-c), the proto-forms were reconstructed with a change in the [ATR] feature from [+ATR] vowels from the Portuguese, resulting in [-ATR] vowels in the reconstructed forms. This can also be observed in the reflexes of cognates, in which, in some daughter languages—such as CRI ['ɔzɛ] and AOA ['ɔðɛ]—the contemporary lexical item has [-ATR] vowels, whereas other daughter languages have [+ATR] vowels, such as FAB ['oze] and PRE ['oze] (cf. (7-ci)). As usual, the trigger for the process is the stressed vowel.

- (7) a. *ɔ...ʼɛ < [o...ʼɛ]
 (i) *ɔ'pɛ⁵ < o pé 'foot'
 (ii) *ɔ'sɛ < o céu 'sky'
 (iii) *tlɔ'mɛNtu 'mess' < tormento
 (iv) *saN'tɔmɛ 'San Thome' < Santomé
- b. *'ɔ...ɛ < ['ɔ... e]
 (i) *'nɔvɛ < nove 'nine'
 (ii) *'ɔmɛ < homem 'man'
 (iii) *aN'tɔNtɛ 'the day before yesterday' < anteontem
- c. *'ɔ...ɛ < ['o...e]
 (i) *'ɔzɛ 'today' < hoje
 (ii) *ɔ'rɛja 'ear' < orelha

⁵ In some examples, there is a process of agglutination of the Portuguese etymon, irrelevant to the discussion presented here (see LADHAMS, 2007; HAGEMEIJER, 2009). This incorporation into nominal words occurred historically, but it was more common in Lung'Ie.

A similar pattern of agreement can be observed in the group of proto-forms whose result was $*\text{'}\text{ɔ}...\text{ɔ}$ from sequences of Portuguese etyma $[\text{o}...\text{'}\text{ɔ}]$ in (8-a), and $[\text{'}\text{ɔ}...\text{o}]$ in (8-b). The presence of a $[-\text{ATR}]$ mid-vowel in (8-a) and (8-b) may have triggered harmony. In addition, the data in (8-c) consistently show the final vowel as a target for the application of a progressive harmonic process; in other words, the final non-stressed vowel is replaced by $[\text{ɔ}]$. This would not be a problem because $[\text{a}]$ is a $[-\text{ATR}]$ vowel (therefore having the same $[-\text{ATR}]$ feature quality). However, $[\text{a}]$ has functioned as a neutral element to the application of $[\text{ATR}]$ harmony elsewhere (see NEVINS, 2010). The proto-form in (8-ci) was reconstructed with $[-\text{ATR}]$ vowels, which is supported by the data from three daughter languages, except for FAB, which has the reflex $[\text{'xobolo}]$ with $[\text{+ATR}]$ vowels.

- (8) a. $*\text{'}\text{ɔ}...\text{ɔ} < [\text{o}...\text{'}\text{ɔ}]$
 (i) $*\text{'}\text{ɔ}'\text{p}\text{ɔ}$ $\text{'powder}' < \text{o pó}$
 b. $*\text{'}\text{ɔ}...\text{ɔ} < [\text{'}\text{ɔ}...\text{o}]$
 (i) $*\text{'}\text{b}\text{ɔ}'\text{d}\text{ɔ}$ $\text{'pier}' < \text{bordo}$
 (ii) $*\text{'}\text{ɔ}'\text{s}\text{ɔ}$ $\text{'bone}' < \text{osso}$
 c. $*\text{'}\text{ɔ}...\text{ɔ} < [\text{'}\text{ɔ}...\text{a}]$
 (i) $*\text{'}\text{k}\text{ɔ}'\text{b}\text{r}\text{ɔ}$ $\text{'snake}' < \text{cobra}$
 (ii) $*\text{'}\text{k}\text{ɔ}'\text{d}\text{ɔ}$ $\text{'rope}' < \text{corda}$
 (iii) $*\text{'}\text{p}\text{ɔ}'\text{t}\text{ɔ}$ $\text{'door}' < \text{porta}$

The proto-forms in (9a), with $*\text{'}\text{ɛ}...\text{ɛ}$, emerged from the sequences $[\text{'}\text{ɛ}...\text{e}]$ and $[\text{'}\text{e}...\text{e}]$. In (9-a), the $[\text{+ATR}]$ vowel in the post-tonic syllable is the target for harmony triggered by the $[-\text{ATR}]$ vowel in the stressed syllable. However, in (9-b), there was an unmotivated change in the $[\text{ATR}]$ feature of the Portuguese etymon $[\text{e}]$, although the presence of the nasal coda (in the examples (9-bi), (9-b-ii) and (10-a-ii)) may have played a role as a triggering element, insofar as it was interpreted as a factor favouring the feature $[-\text{ATR}]$. In modern

European Portuguese, nasalised mid-vowels are [-ATR] (see WETZELS, 1997).

- (9) a. *'ε...ε < ['ε...e]
 (i) *'fεblε < *febre* 'fever'
 (ii) *'sεtε < *sete* 'seven'
- b. *'ε...ε < ['eN...e]
 (i) *'sεNplε < *sempre* 'always'
 (ii) *sε'tεNblu < *Setembro* 'September'

In the proto-form *vlε'gɔɲa in (10-a), both [+ATR] vowels of the Portuguese etymon had their value changed to [-ATR]. Again, the trigger is the stressed vowel, in this case, a nasal coda. Nevertheless, this example reveals that although there was already an agreement of [+ATR] features in the Portuguese etymon, the change in PGG remained in the daughter languages when reflexes were observed in (10-a-ii). Fa d'Ambô was the only daughter language to keep [+ATR] vowels, but with a metathesis between the two initial syllables and a change from [e] to [o] in the syllable [ver], as in (10-a-iii).

- (10) a. * ε...'ɔ < [e... 'oN] and reflexes
 (i) *vlε'gɔɲa 'shame' < *vergonha*
 See:
 (ii) CRI [vlε'gɔɲɐ], PRE [vɔ'gjɐ], AOA [ve'gɔɲɐ]
 (iii) FAB [go'vojɐ]

The example in (11) also shows changes in the quality of the [ATR] feature.

- (11) a. *'ɔ...ε < [o...a...o]
 (i) *rɔ'vwε < *orvalho* 'dew'

However, the complexity of the reflections in CRI [lɔ'vɛ], PRE [rɔ'vwɛ] and AOA [la'vwe] suggests a path involving other processes from the Portuguese etymon to the forms in daughter languages, such as metathesis and resolution of diphthongs. The item in FAB, [se'lema], is not a cognate and has thus been excluded.

Finally, the data in (12) have sequences of syllables with harmonic vowels in two items: in (12-a), in a context in which the insertion of a vowel copy was expected, and in (12-b), arising from a change from anterior mid-vowels to posterior mid-vowels. Therefore, such are the examples whose genesis remains obscure, but they maintain the tendency to avoid discordant mid-vowels in relation to the [ATR] feature. In (12), however, one cannot argue for the presence of parasitic harmony due to the disagreement of the vowel features.

- (12) a. *'o...e < [a...'or#]
 (i) *a'mole 'love' < *amor*
 b. *o...'o...u < ; [[o][o...'e...o]]
 (i) *o'zoju 'knee' < *joelho*

2.4 Non-harmonic forms

In the *corpus*, 47 items (of 536) can be classified as non-harmonic if we consider that in the Portuguese etymon, there was a sequence of two vowels in contiguous syllables, one being [aATR] and the other [βATR], as in ['mɔpe], in which the first vowel is [-ATR] and the second is [+ATR]. Non-harmonic items represent about 12% (47 items) of all non-verbal elements of the *corpus* and a slightly higher number than all harmonic data, which totalled 46 items with [aATR] mid-vowels in contiguous syllables. Therefore, the question of the existence of PGG proto-forms in which VH did not apply, although there was a favourable context, cannot be ignored. For example, the proto-form *'tɔsi 'cough', has four reflexes in the daughter languages with the final vowel [i],

although the Portuguese etymon has a final [e]: CRI ['tɔʃĩ], FAB ['tɔʃĩ], PRE ['tɔʃĩ] and AOA ['tɔsi]. The proto-form *'pɛli 'skin' in (13-b) is another case of disharmony in a favourable context, confirmed by the reflexes in CRI ['pɛli], FAB ['pɛli], PRE ['pɛli] and AOA ['pɛli]. Finally, the proto-form *ka'belu 'hair' in (13-c) contains reflexes in [u], although it did not keep the final vowel [o] in any form of the daughter languages. Nonetheless, in FAB and in PRE, the quality of stressed [e] was changed to [ɛ]: CRI [ka'belu], FAB [xa'bɛlu], PRE [ka'bɛlʊ] and AOA [ka'belu]. In (13-d), for example, the proto-form contains a final disharmonic vowel in relation to the vowel of the stressed syllable, whereas the Portuguese etymon contains two [+ATR] vowels; therefore, it is already harmonic. In this example, there was once again a change in the quality of the [ATR] feature if we consider the Portuguese etymon. However, the reflexes in the daughter languages (CRI ['ɔvu], FAB ['ɔvu], PRE ['ɔvu] and AOA ['ɔvu]) consistently support the proposition of the proto-form *'ɔvu.

- (13) a. 'ɔ...i < ['ɔ...e]
 (i) *'tɔsi 'cough' < *tosse*
 b. *'ɛ...i < ['ɛ...e]
 (i) *'pɛli 'skin' < *pele*
 c. *'e...u < ['e...o]
 (i) *ka'belu 'hair' < *cabelo*
 d. *'ɔ...u < ['o...o]
 (i) *'ɔvu 'egg' < *ovo*

The issue of the neutralisation of mid-vowels <e, o> in the final unstressed position in 16th-century Portuguese is controversial. Naro (1973) states that in the 16th century and in the archaic period, there was no complete raising of <e> and <o> to [i] and [u] in the final stressed position. In turn, Teyssier (2001) argues that there is no evidence to support the pronunciation of <e> and <o> as [i] and [u],

respectively. However, Maia (1986) argues that there was phonetic variation and that <e> and <o> could be pronounced as [i] and [e] and [u] and [o], respectively, in the final non-stressed position. Fonte (2017, p. 194) shows that

‘...the rising of the final unstressed vowel, although it might have occurred in Galician-Portuguese, middle Portuguese and early modern Portuguese, was by no means preponderant in the language until the second half of the 16th century.’

Consequently, it is reasonable to argue that at least some forms would have entered PGG already with the final non-stressed high vowels. However, the pattern presented by cognates of daughter languages and proto-forms cannot be attributed solely to a late influence of Portuguese or even Spanish (in the case of FAB); a simultaneous application to the same cognates would be unlikely in all daughter languages, and harmonic forms, to which the neutralisation rule can also be applied, were not influenced by the colonial languages. However, this is difficult to prove without further synchronic data on the variation of final non-stressed vowels in those languages. For example, PRE currently has a variation in the final non-stressed vowel in harmonic form, as in /ɔNzɛ/ [ʔzɛ] ~ [ʔzi] ~ [ʔzi] *onze* ‘eleven’, /ʒipe/ [ʒipe] ~ [ʒipi] *jipe* ‘jeep’, /fologo/ [fologo] ~ [fologu] *fôlego* ‘breath’ (AGOSTINHO, 2016).

2.5 Compounds

As far as compounds are concerned, the vowel harmony process does not occur even in contexts of [ATR] mid-vowels in contiguous syllables. In (14), the compound proto-form *b_εga ko'le *diarrhoea*, formed by *b_εga *barriga* ‘belly’ and *ko'le *correr* ‘run’, has in the item *b_εga a stressed [-ATR] mid-vowel, [ɛ], and in the second element of

the compound, the proto-form has [+ATR] vowels, [o] and [e], the latter stressed. Regarding the possibility of the harmony being bidirectional, it is not clear from example (14) whether the process would be triggered by the stressed vowels [ɛ] or [e].

- (14) *'bɛga kole > (('bɛga) (ko'le)) *diarréia* 'diarrhoea'
 [['bɛga][ko'le]] (CRI, AOA) 'diarrhoea'
 ea'

Nevertheless, the data in (14) show that, at least in compounds, when two prosodic words form a new lexical item, the obligatory agreement of [ATR] features of vowels in adjacent syllables occurs in the domain of each prosodic word, thus avoiding that *'bɛga goes to 'bega, or that *ko'le goes to kɔ'le. This can also be observed in (16), preventing the compound *(('sejta)('fɛla)) from being (('sejta)('fela)) with the [-ATR] feature of the vowel [ɛ] functioning as the harmonising element or *((('sejta)('fɛla)), in which the [+ATR] vowel of the first component, 'sejta, would trigger a change in the second component. However, the change in the quality of the [ATR] feature in the stressed vowel in the component *'fɛla cannot be motivated by any element in the compound because in the Portuguese etymon, there was the [+ATR] vowel, as in 'fela. Therefore, this change cannot be associated with the regularity of a process whose target was the Portuguese etymon.

- (15) *'sejta 'fɛla 'Friday' > *sexta-feira*
 [['sejʔɐ]['fɛʔɐ]] (CRI) 'Friday'
 [['θeta]['fɛla]] (AOA) 'Friday'

In addition, there are examples in compounds to which harmony does not apply. This blockage is again associated with the prosodic structure of the compound. In PGG, the proto-form *bwe *boi* 'ox' was reconstructed from cognates of the four languages: in the CRI and FAB,

from ['bwe] and in PRE and AOA from [[u'bwe] ['ɔ̃mi]] and [['bwe] ['ɔ̃mɛ]] *boi macho* 'ox' (literally *male ox*), respectively. In the case of PRE, the contiguous vowels [e] and [ɔ̃] in [[u'bwe] ['ɔ̃mi]] are not harmonic, whereas in AOA, the same non-harmonic double also occurs. The *vh* subsists within the second component of the compound, ['ɔ̃mɛ]. Therefore, sequences of two words in the PGG and in the daughter languages, despite constituting a semantic unit, must be considered post-lexical compounds, that is, formed after a cycle of the application of phonological rules (KIPARSKY, 1982). In other words, vowel harmony can be restricted to the root of the word, considering only the prosodic word as the phonological unity of the domain, as suggested by the data in (15), and excluding post-lexical compounds, because they are treated as two prosodic words.

2.6 *Non-Portuguese etyma*

There are, in addition to the Portuguese etyma in the *corpus*, nominal examples of non-Portuguese origin in (16). In these examples, it cannot be determined whether the pattern of agreement with the [ATR] feature has been incorporated as in substrate languages or whether it is an incorporation following the vowel copy rule. Although, *bo'bo is the only proto-form that has reflexes in the four languages because among others, (16-b) and (16-c) have cognates in CRI, PRE and AOA, (16-d) in CRI and FAB and (16-e) in CRI, FAB and AOA.

(16) Non-Portuguese origin

- a. *bo'bo 'ripe' < ?bobo⁶
- b. *ake're 'toad' < ?akere⁷
- c. *'tjɔkɔ 'little' < ?
- d. *klɔ'klɔsɔ 'neck' < ?gogongo⁸
- e. *'kɔbɔ 'hole' < ?⁹

In summary, the occurrence of non-Portuguese etyma brings a corroborative component to the discussion because they eliminate Portuguese etyma as unique objects of the process. At the same time, Portuguese, as a language in which there is no [ATR] vowel harmony, functioned as a destabilising element of the incipient system, which may have proved crucial in the early years when speakers implemented harmony when there was a context (i.e., the presence of a mid-vowel in the tonic syllable). However, not all items susceptible to the process were targeted (which resulted in forms in which not all vowels were in the contiguous syllables involved) and, finally, the [ATR] features were changed either for [+ATR] or [-ATR], both in PGG and in the formation of daughter languages. Nevertheless, it would be unreasonable to attribute the harmonic features of the four daughter languages to chance and to individual developments; proper rules of harmony in each language would hardly achieve the same results of the sets of cognates without a common origin.

⁶ Ladhams (2007: 14) states that this lexical item may be of Kikongo origin from the etymon *booba*. Maurer (2009), in turn, suggests that the origin may be Kikongo, *bôba* 'be red, ripe, well cooked', or edo, *boô!* 'ripen, be ripe'.

⁷ Ladhams (2007: 14) states that this lexical item may be of Yoruba origin from the etymon *akere*. Maurer (2009) suggests the Edo form *ɛkɪɛ* 'frog' or the Yoruba *ākere* 'a striped frog with smooth skin'.

⁸ Ladhams (2007: 14) states that this lexical item may be of Yoruba origin from the etymon *gogongo*.

⁹ In CRI and AOA: ['kɔbɔ], on FAB: ['xɔbɔ], and in PRE, a non-cognate form: [ɔkɛ'rɛ].

3. Final remarks

The number of proto-forms in which [aATR] harmony seems not to be applied in a favourable context and the maintenance of such non-harmonic lexical items in daughter languages, as shown in (13), create challenges for a unified analysis of the vowel harmony of PGG. The reason for the non-application of the process suggests that the neutralisation of final vowels was a process robust enough to be incorporated by proto-Creole speakers at the beginning of the colonisation of the São Tomé Island. In addition, the maintenance of proto-forms in daughter languages can be better understood considering the inadequacy of the context, because the presence of a non-mid-vowel in such proto-forms would block the application of harmony in the PGG. At the same time, proto-forms show that there could be a kind of harmony of contrastive [ATR] mid-vowels in contiguous syllables, but the process was limited and should be separated from the vowel copy process:

(17) a. Harmony [aATR]

- (i) compulsory agreement of the [ATR] feature in mid-vowels in contiguous syllables;
- (ii) maintenance of [aATR] mid-vowels of the Portuguese etymon;
- (iii) maintenance of [aatr] mid-vowels of the non-Portuguese etymon.

b. Vowel copy (harmonic)

- (i) insertion of a vowel copy in Portuguese etyma in order to avoid roots ending in consonants.

In summary, [aATR] mid-vowel sequences in contiguous syllables in (17-a-i) were kept in the PGG proto-forms when already present in the Portuguese etymon. Accordingly, it was a case of correspondence to

the original form, although convergent with the harmonic pattern of [aATR] mid-vowels. However, daughter languages have not always remained faithful to such proto-forms because the dynamics of linguistic contact and internal processes may have caused some *ad hoc* outputs to some lexical items. The type of harmony described in (17-b-i) is not limited to mid-vowels. It involves all vowels and thus diverges from the other types. The generalisation pointed out in (17-a-ii) suggests a similarity with (17-a-i), but this can only be verified via further studies on substrate languages. At the same time, proto-forms resulting from this process generated harmonic reflexes in the daughter languages of PGG.

The compulsory agreement between mid-vowels in contiguous syllables regarding the [ATR] feature (see (17-a-iii)), is robust in PGG. On the one hand, in 16th-century Portuguese, [-ATR] mid-vowels occurred only in a stressed position, and thus the agreement between [-ATR] mid-vowels in contiguous syllables with a Portuguese etymon in PGG is a generalisation from the stressed [-ATR], a classic example of [ATR] vowel harmony. On the other hand, some proto-forms with [-ATR] mid-vowels originated from Portuguese etyma with [+ATR] mid-vowels. Thus, speakers in the early stages of PGG, possibly influenced by [ATR] mid-vowel harmony patterns of some of their native languages, applied generalisations both in favouring contexts—that is, with a [-ATR] mid-vowel in a stressed position—and in unfavourable contexts—that is, with [+ATR] mid-vowels in a stressed position, interpreted as [-ATR]. The few cases of harmony with [+ATR] mid-vowels, as described in (12), are examples of unfavourable contexts, but the results indicate that mid-vowels kept their similar features.

Therefore, the representation in (18) summarises the possibilities indicated by CV.CV word sequences with [ATR] features. The agrammatical nature of the examples in (18-b) is the result of sequences of mid-vowels in contiguous syllables with discordant [ATR] features.

- (18) a) CV_[-ATR].CV_[-ATR]
 CV_[+ATR].CV_[+ATR]
 b) ~~CV_[-ATR].CV_[+ATR]~~
 ~~CV_[+ATR].CV_[-ATR]~~

Here, we showed that there are more cases in which there was no application of the process than harmonic cases, based on the proto-forms and on patterns of harmonisation with [ATR] mid-vowels in contiguous syllables in the PGG. In addition, we argued that there were processes of vowel harmony in the proto-Creole of the Gulf of Guinea active in the nativisation of Portuguese and non-Portuguese etyma. However, it was necessary to classify the processes basically into two types: (i) those presenting pseudo-harmony (i.e., items in which vowels were copied or a harmonic pattern was maintained in the PGG from a past synchrony) and (ii) those presenting a real harmony (i.e., items in which the application of a harmony rule based on mid-vowels with identical [ATR] features is defensible). Finally, examples in which harmony does not apply internally to the root and the establishment of the prosodic word as the domain of the process are contributions that shed light on the analysis of the process in modern Santome, Angolar, Lung'Ie and Fa d'Ambô. Thus, the analysis of our data suggests that [ATR] mid-vowel harmony in contiguous syllables in PGG was unstable.

References

- AGOSTINHO, A. L. S. **Fonologia do Lung'Ie**. Muenchen: Lincom, 2016.
- ARAUJO, G. A.; BANDEIRA, M. **Dicionário Angolar-Português**. São Paulo: FFLCH, on press.
- ARAUJO, G. A.; HAGEMEIJER, T. **Dicionário Santome-Português / Português-Santome**. São Paulo: Hedra, 2013.

BALDUINO, A. M.; AGOSTINHO, A. L. S.; ARAUJO, G. A., and CHRISTOFOLETTI, A. S. A nasalidade vocálica em santome e lung'ie. **Papia**, São Paulo, n. 25, p. 7-25, 2015.

BANDEIRA, M. **Reconstrução fonológica e lexical do Protocrioulo do Golfo da Guiné**. Muenchen: Lincom, 2017.

BANDEIRA, Manuele; ARAUJO, Gabriel Antunes de; FINBOW Thomas. Liquids in the Proto-Creole of the Gulf of Guinea and in its daughter languages. USP, draft, 2019.

BECKMAN, J. **Positional Faithfulness**. Unpublished thesis (PhD), University of Massachusetts, 1998.

COLE, J; TRIGO, L. Parasitic harmony. In: VAN DER HULST, H. **Features, segmental structure, and harmony processes II**. Dordrecht: Foris, 1988, p. 19-38.

FERRAZ, L. I. **The Creole of S. Thomé**. Johannesburg: Witwatersrand University Press, 1979.

FONTE, J. S. O vocalismo átono na história da língua portuguesa. **Alfa**, São Paulo, n. 61, p. 169-199, 2017. DOI: 169-199. doi.org/10.1590/1981-5794-1704-7.

HAGEMEIJER, T. Initial vowel agglutination in the Gulf of Guinea creoles. In: ABOH E., and SMITH N. **Complex processes in new languages**. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2009, p. 29-50. DOI: <https://doi.org/10.10>

LADHAMS, J. Article Agglutination and the African Contribution to the Portuguese-based Creoles, 2007. Disponível em: <goo.gl/xBYH3F> Acesso em: 10 jan. 2018.

KIPARSKY, P. Lexical Morphology and Phonology. In: YANG, I. S. **Linguistics in the Morning Calm**. Hanshin: Seoul, 1982, p. 3-91.

MAIA, C. A. **História do galego português. Estado linguístico da Galiza e do Noroeste de Portugal desde o século XIII ao século XVI. (Com referência à situação do galego moderno)**. Coimbra: Instituto Nacional de Investigação Científica, 1986.

MAURER, P. **L'Angolar: Un créole afroportugais parlé à São Tomé; Notes de grammaire, textes, vocabulaires.** Hamburg: Helmut Buske Verlag, 1995.

MAURER, P. **Principense. Grammar, texts, and vocabulary of the Afro-Portuguese creole of the Island of Príncipe, Gulf of Guinea.** London: Battlebridge Publications, 2009.

MOHANAN, P. J. Vowel Harmony. In: EID, M.; ELGIBALI, A.; VERSTEEGH, K.; WOIDICH, M. & ZABORSKI, A. **The Encyclopedia of Arabic Language and Linguistics.** Leiden: Brill Academic Publishing, 2009, p. 676-677. DOI: http://dx.doi.org/10.1163/1570-6699_eall_EALL_COM_0373

NARO, A. A história do e e do o em português: um estudo de deriva linguística. **Estudos diacrônicos**, Petrópolis, p. 9-31, 1973.

NEVINS, A. **Locality in Vowel Harmony.** Cambridge: MIT Press, 2010. DOI: [doi.org/10.7551/mitpress/9780262140973.001.0001](http://dx.doi.org/10.7551/mitpress/9780262140973.001.0001)

O'KEEFE, M. **Akan vowel harmony.** Unpublished Bachelor's thesis. Pennsylvania: Swarthmore College, 2003.

OLA ORIE, O. An alignment-based account of vowel harmony in Ife Yoruba. **Journal of African Languages and Linguistics**, Alemanha, v. 22, p. 117-143, 2001. DOI: <https://doi.org/10.1515/jall.2001.001>

PULLEYBLANK, G. D. Vocalic underspecification in Yoruba. **Linguistic Inquiry**, Massachusetts, v. 19, p. 233-270, 1988.

SEGORBE ZAMORA, A. **Gramática descriptiva del fa d'ambô.** Barcelona: CEIBA, 2010.

TEYSSIER, P. **História da língua portuguesa.** 2ª ed. São Paulo: Martins Fontes, 2011.

VIARO, M. E. **Etimologia.** São Paulo: Contexto, 2011.

WETZELS, L. The lexical representation of nasality in Brazilian Portuguese. **Probus** v. 9, p. 203-232, 1997. DOI: <https://doi.org/10.1515/prbs.1997.9.2.203>

