

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS  
DEPARTAMENTO DE LETRAS CLÁSSICAS E VERNÁCULAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOLOGIA E LÍNGUA PORTUGUESA

MARCUS VINÍCIUS KNUPP BARRETTO

**Contribuições da língua portuguesa e das línguas africanas  
quicongo e bini na constituição do crioulo são-tomense**

São Paulo  
2008

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS  
DEPARTAMENTO DE LETRAS CLÁSSICAS E VERNÁCULAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOLOGIA E LÍNGUA PORTUGUESA

**Contribuições da língua portuguesa e das línguas africanas  
quicongo e bini na constituição do crioulo são-tomense**

Marcus Vinícius Knupp Barretto

Dissertação apresentada ao Programa de Pós-graduação em Filologia e Língua Portuguesa do Departamento de Letras Clássicas e Vernáculas da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, para a obtenção do título de mestre em Letras.

Orientador: Prof. Dr. Gabriel Antunes de Araújo

São Paulo  
2008

## AGRADECIMENTOS

“Nunya, adidoe, asi metunee o<sup>1</sup>.”

Provérbio africano ewe e akan

Meus agradecimentos a todos que me ajudaram a abraçar o conhecimento.

Ao professor Dr. Gabriel Antunes de Araújo, pela orientação, pelas aulas no curso de crioulistica, por todo o auxílio durante o mestrado, além de toda a confiança que tem depositado em mim.

À professora Dr<sup>a</sup>. Maria Margarida Taddoni Petter, pelas aulas no curso de línguas africanas, pelo auxílio durante o mestrado e todas as sugestões dadas para a dissertação tanto na qualificação quanto na defesa. Ao professor Dr. Manoel Mourivaldo Santiago Almeida, pelas norteadoras sugestões no exame de qualificação. Ao professor Dr. Rogério Vicente Ferreira, pelas sugestões na defesa. Ao professor Dr. Didier Demolin, por compartilhar seus notáveis conhecimentos de gramática de línguas africanas. Ao professor Dr. Ataliba Teixeira de Castilho, por me permitir ser aluno de sua última turma antes de sua aposentadoria, possibilitando que eu absorvesse parte de sua erudição. À professora Dr<sup>a</sup>. Márcia Duarte de Oliveira, pelas aulas no curso de crioulistica, pelo auxílio e orientação. Aos professores Dr. Nicolas Quint e Dr. Jean-Louis Rougé de quem fui efêmero aluno, nos minicursos e palestras oferecidos no Brasil.

À Universidade de São Paulo, aos colegas dos cursos da USP, a todos os amigos que fiz em São Paulo, durante minha breve estada nessa cidade, e que de alguma forma me ajudaram.

A todos os cientistas que não conheci pessoalmente e não poderei conhecer, mas cujas obras serviram de base teórica para minha dissertação. A todos esses que abriram o caminho facilitando meu trabalho.

À minha família, em especial à minha mãe Neuma Knupp Barretto, com quem sempre pude contar, mesmo nos momentos mais difíceis.

---

<sup>1</sup> O conhecimento é como o tronco de um baobá, ninguém pode abraçá-lo sozinho.

## RESUMO

O objetivo desta dissertação é apresentar e discutir alguns processos fonológicos de adição e subtração de elementos (metaplasmos) na língua sãotomense. Neste trabalho, faremos uma comparação entre as contribuições das línguas portuguesa, quicongo e bini.

Entre os séculos XV e XVI, diversas línguas nasceram do contato entre europeus e povos da África, Ásia e América. Chamadas de pidgins e crioulos, essas línguas contam com contribuições linguísticas da língua do povo dominador (língua de superstrato) e com contribuições da(s) língua(s) do(s) povo(s) dominado(s) (língua(s) de substrato).

O sãotomense, língua falada atualmente na República de São Tomé e Príncipe, é uma dessas línguas, classificada como crioulo de base portuguesa, e conta com o português seiscentista como língua de superstrato e com línguas africanas, dentre elas o quicongo e o bini como línguas de substrato. Ao longo deste trabalho, analisaremos algumas das influências das línguas de substrato e superstrato na constituição do sãotomense.

As contribuições das línguas de superstrato estão, majoritariamente, relacionadas à composição do léxico e as das línguas de substrato na fonologia, morfologia e sintaxe, embora também haja traços inovadores. No caso do sãotomense, as palavras portuguesas, ao entrarem no léxico do sãotomense, sofreram metaplasmos para se adequar à estrutura das línguas africanas dos primeiros falantes, sem, contudo, evitar que a língua portuguesa também contribuísse para a constituição da fonologia do sãotomense. Uma das contribuições do quicongo na fonologia do sãotomense é o lambdacismo — transformação de [r] em [l] durante o processo de empréstimos, enquanto a língua portuguesa contribuiu com a eliminação do sistema tonal, presente em quicongo e bini, mas não em português.

**Palavras-chave:** contato de línguas — crioulos — sãotomense — substrato — superstrato — metaplasmos

## ABSTRACT

The goal of this dissertation is to describe and analyse some phonological aspects of Sãotomense. In this word, we compare a number of linguistics contributions from the Portuguese, Kongo and Bini languages to Sãotomense, a Portuguese-based Creole spoken in the Democratic Republic of São Tomé and Príncipe.

A Creole language displays linguistic characteristic both from its superstratum and its substratum languages. Sãotomense has the seventeenth century's Portuguese as its superstratum language and many African languages, among them Kongo and Bini, as its substrata languages.

In this work, I intend to analyze some of the influences of these strata languages in the formation of Sãotomense phonology. In general terms, most of the contributions from the superstratum languages are related to the Lexicon. Substratum languages, by its turn, heavily contribute to the phonology, morphology and syntax, although there are in the Creoles languages innovative linguistics aspects as well. In the specific case of Sãotomense, Portuguese words undergone many linguistics processes, some of them called metaplasms, in order to be adapted by the structure of African languages speakers, but this fact did not avoid that Portuguese language also contributed to the phonology constitution of Sãotomense. A possible African contribution to the phonology of Sãotomense is the so-called lambdacism — the transformation of a [r] into a [l] during the process of loanword adaptation from the Kongo language. Portuguese, for example, probably, contributed with the elimination of tones, present in Kongo and Bini and in many others African languages, but not in Portuguese.

**Key words:** languages in contact — creoles — sãotomense — substratum — superstratum — metaplasm

## LISTA DE MAPAS

MAPA 1 .....	p. 15
MAPA 2 .....	p. 16
MAPA 3 .....	p. 19

## LISTA DE TABELAS

TABELA 1 .....	p. 17
TABELA 2 .....	p. 19
TABELA 3 .....	p. 36
TABELA 4 .....	p. 36
TABELA 5 .....	p. 41
TABELA 6 .....	p. 43
TABELA 7 .....	p. 45
TABELA 8 .....	p. 48
TABELA 9 .....	p. 49
TABELA 10 .....	p. 49
TABELA 11 .....	p. 55
TABELA 12 .....	p. 55
TABELA 13 .....	p. 58
TABELA 14 .....	p. 61
TABELA 15 .....	p. 61
TABELA 16 .....	p. 66
TABELA 17 .....	p. 66
TABELA 18 .....	p. 66
TABELA 19 .....	p. 71
TABELA 20 .....	p. 72
TABELA 21 .....	p. 75
TABELA 22 .....	p. 77
TABELA 23 .....	p. 83
TABELA 24 .....	p. 97
TABELA 25 .....	p. 108
TABELA 26 .....	p. 108
TABELA 27 .....	p. 113
TABELA 28 .....	p. 115

## Sumário

1. Introdução .....	3
1.1. Sobre o propósito deste trabalho .....	3
1.2. Apresentação .....	3
1.3. Metodologia .....	7
1.4. Geografia de São Tomé e Príncipe .....	9
1.4.1. Geografia física.....	9
1.4.2. Geografia humana .....	10
1.4.3. Geografia linguística .....	11
1.5. História de São Tomé e Príncipe .....	13
2. O contato linguístico .....	16
2.1. O contato interlinguístico .....	16
2.2. O processo de aquisição da linguagem .....	17
2.2.1. O processo monitorado de aquisição de uma língua estrangeira .....	17
2.2.2. O processo não-monitorado de aquisição de uma língua estrangeira .....	18
2.3. Pidgin .....	20
2.4. Crioulo .....	20
2.5. Substrato e superstrato .....	22
2.6. Os crioulos e a evolução das línguas .....	23
2.7. Os crioulos e a taxonomia linguística .....	25
2.7.1. A classificação genética .....	28
2.8. Hipóteses para o surgimento dos pidgins e crioulos .....	29
2.8.1. A teoria monogenética .....	29
2.8.2. Baby-talk .....	31
2.8.3. Bioprograma .....	32
2.8.4. Gíria de marinheiros .....	32
2.8.5. Conclusão .....	33
2.9. O problema da deriva linguística .....	33
3. Os substratos africanos dos crioulos do Golfo da Guiné .....	35
3.1. Quicongo .....	36
3.2. Bini .....	41
4. O superstrato português dos crioulos do Golfo da Guiné .....	45
4.1. Características do português arcaico .....	46
4.1.1. Fonética .....	46
4.1.2. Morfologia .....	47
4.1.3. Sintaxe .....	47
4.2. Algumas características do português seiscentista .....	47
4.2.1. Fonologia .....	48
4.2.1.1. O sistema consonantal .....	48
4.2.1.2. As sibilantes .....	49
4.2.1.3. Os fones [ʒ] e [ʃ] .....	50
4.2.1.4. A flutuação entre [b] e [v] .....	51
4.2.1.5. Inventário fonológico das consoantes .....	52
4.2.1.6. O sistema vocálico .....	54
4.2.1.7. O ditongo [ɐũ] .....	55
4.2.1.8. O ditongo [ow] .....	56
4.2.1.9. O ditongo [ej] .....	57
5. Fonética e fonologia do sãotomense .....	59
5.1. Inventário fonético das consoantes .....	59
5.2. Inventário fonológico das consoantes .....	63
5.3. Inventário fonético das vogais .....	64
5.4. Inventário fonológico das vogais .....	65
5.5. A estrutura silábica .....	67
6. Metaplasmos .....	72

6.1. A estrutura das sílabas e as regras fonotáticas .....	73
6.2. Metaplasmos por permuta .....	74
6.2.1. Nasalização ou nasalização .....	74
6.2.2. Palatalização .....	75
6.2.3. Despalatalização .....	76
6.2.4. Lambdacismo .....	77
6.2.5. Ieísmo .....	78
6.3. Metaplasmos por aumento .....	79
6.3.1. A prótese .....	79
6.3.2. A epêntese .....	79
6.3.2.1. A ditongação .....	80
6.3.3. A paragoge ou epítese .....	80
6.4. Metaplasmos por subtração .....	81
6.4.1. A aférese .....	81
6.4.2. A síncope .....	82
6.4.2.1. A monotongação .....	82
6.4.3. A apócope .....	83
6.4.4. A sinalefa ou elisão .....	83
6.5. Metaplasmos por transposição .....	83
6.5.1. A metátese .....	84
6.5.2. Hiperbibasmo .....	84
7. Outras contribuições do substrato .....	85
7.1. Estrutura silábica do sãotomense e do principense em comparação com as estruturas silábicas das línguas negro-africanas .....	85
7.2. Influência da estrutura silábica africana no principense .....	87
7.3. Influências fonológicas africanas no sãotomense .....	89
7.3.1. O desaparecimento da vibrante múltipla /r/ e do tepe /r/ .....	89
7.4. Influências sintáticas .....	91
7.4.1. Pronomes disjuntivos .....	91
7.4.2. Negação disjuntiva .....	92
7.5. Influências lexicais e semânticas do quicongo e do bini no sãotomense .....	93
7.6. Os ideofones .....	95
8. Outras contribuições do superstrato .....	98
8.1. Fonologia .....	98
8.2. Morfologia .....	98
9. Contribuições conjuntas e elementos autênticos .....	100
9.1. Verbos.....	101
9.2. Numerais.....	106
10. Considerações finais .....	109
11. Referências bibliográficas .....	111



## 1. INTRODUÇÃO

### 1.1. Sobre o propósito deste trabalho

A presente dissertação trata das contribuições da língua portuguesa e das línguas africanas quicongo e bini para a constituição do crioulo de base portuguesa sãotomense (também conhecido como forro, 'lungwa san'tome e 'lungwa di 'tela).

Neste trabalho, faremos uma comparação entre a língua portuguesa, o sãotomense, o quicongo e o bini. Como o português e as línguas africanas quicongo e bini são línguas de troncos diferentes, não serão feitas comparações diretas entre elas, mas comparações do português, do quicongo e do bini com o sãotomense, língua crioula que teve o português como superstrato (língua lexificadora) e o quicongo e o bini como substrato (línguas que influenciaram a fonologia, a morfologia, a sintaxe e a semântica).

Em relação às línguas africanas, este trabalho se propõe a demonstrar as influências da fonologia na constituição do léxico sãotomense que, tendo sido composto com palavras da língua portuguesa, sofreu alterações fonéticas durante seu empréstimo.

Em relação à língua portuguesa, este trabalho se propõe a demonstrar algumas contribuições ao crioulo sãotomense além do léxico, ou seja, contribuições portuguesas à fonologia, morfologia e sintaxe do sãotomense.

Será tomada uma atitude comparativa durante toda a dissertação, tentando rastrear, quando possível, as contribuições portuguesas ou africanas no crioulo estudado.

### 1.2. Apresentação

Quase todas as línguas possuem, em maior ou menor grau, vocábulos emprestados de outras línguas. O português atual possui cerca de sete mil palavras oriundas do árabe<sup>2</sup>, fruto de um contato com muçulmanos do norte da África na Península Ibérica por mais de quinhentos anos (Teyssier, 2004: 5-6)<sup>3</sup>, quando a língua árabe manteve uma posição político-linguística de primazia. Já o português do Brasil conta ainda com mais de 2,5 mil palavras oriundas do tupi<sup>4</sup>, fruto da convivência de quase três séculos entre

---

<sup>2</sup> Contagem do dicionário eletrônico Aurélio.

<sup>3</sup> Segunda edição. A primeira edição francesa é de 1980, e a primeira brasileira é de 1997.

<sup>4</sup> Segundo o dicionário Aurélio (2006), as palavras são listadas como tupis, mas segundo Agostinho e Araújo (2007) “A língua tupi, efetivamente, nunca existiu. A expressão ‘língua tupi’ refere-se ao dialeto unificado dos instrumentos linguísticos empregados pelos jesuítas, sem especificação das diferenças

lusófonos e indígenas da costa brasileira, com os portugueses em posição de primazia; e centenas de palavras de origem africana, devido aos quase três séculos de tráfico de escravos, em que os portugueses mantiveram forte domínio sobre os povos africanos.

Durante o século XIX, milhares de palavras entraram na língua portuguesa<sup>5</sup> e, mais recentemente, com a globalização e posição econômica privilegiada dos países anglófonos, sobretudo os Estados Unidos e o Reino Unido, milhares de outras palavras inglesas entram, como já entraram anteriormente, e continuarão entrando na língua portuguesa, não importando se os falantes das línguas estrangeiras se encontram em posição social, econômica e política superior (como foi com o árabe, francês e inglês) ou inferior (como foi com os indígenas e africanos). Os contatos, ainda, nem sempre precisam ser diretos; brasileiros e portugueses nunca tiveram contato com os astecas do México, mesmo assim, palavras astecas como “tomate”, “chiclete” e “chocolate” entraram e ainda fazem parte do vocabulário cotidiano dos lusófonos.

“Alfinete”, “guaraná”, “caçula”, “manicure” e “bife”, palavras já incorporadas ao idioma português moderno, não causam mais estranheza ao falante nativo de português, que utiliza tais palavras, oriundas, respectivamente, do árabe, tupi, quimbundo, francês e inglês, como palavras legítimas da língua portuguesa (e de fato, o são, devido às adaptações fonológicas, conforme veremos a seguir). Alguns desses vocábulos pouco lembram as palavras de origem, por terem passado pelo processo de adaptação fonológica, semelhante ao que será visto ao longo deste trabalho.

Assim como há centenas de palavras estrangeiras na língua portuguesa, há centenas de palavras portuguesas em dezenas de línguas do mundo.

O Império Português já teve como colônias terras em quase todos os continentes, absorvendo palavras de línguas diversas e espalhando a língua portuguesa entre povos distintos. Ignorando-se a preexistência dos povos autóctones, como se faz ao dizer que Pedro Álvares Cabral descobriu o Brasil (quando grupos indígenas já o habitavam), costuma-se dizer também que os portugueses descobriram o Japão. Durante menos de um século, entre 1571 e 1639, a cidade de Nagazáki esteve entre essas colônias do antigo Império Português, um tempo curto, mas não curto o suficiente para impedir empréstimos lexicais. Segundo Koseki (2007), a colonização portuguesa no Japão legou à língua japonesa mais de quatro mil palavras, destas, mais de quatrocentas são usadas

---

regionais (...). Cientificamente, emprega-se o termo tupi como denominação de um tronco linguístico, embora, nos círculos não-científicos e paracientíficos emprega-se o termo tupi como designação de uma língua indígena brasileira.” (cf. ainda Rodrigues, 1986).

<sup>5</sup> Contagem do dicionário Aurélio eletrônico.

ainda hoje no dia-a-dia da sociedade japonesa, como **pan** “pão”, **shaban** “sabão”, **botan** “botão”, **kirishitan** “cristão” e **arukooru** “álcool”, esta última vinda do árabe para o português, e deste para o japonês. Como visto, as palavras de origem portuguesa usadas no japonês são pronunciadas de forma diferente do português, isso se deve ao processo de adaptação fonológica.

Palavras de origem estrangeira na língua portuguesa, ou da língua portuguesa em línguas estrangeiras, embora sejam contadas às centenas ou aos milhares, ao serem comparadas com o vocabulário total do idioma focado, mostrar-se-ão uma minoria. As quatro mil palavras existentes no português, de origem inglesa, não chegam a 1% do vocabulário total da língua portuguesa, tomando este vocabulário como sendo de 450 mil palavras<sup>6</sup>. O mesmo se diz das palavras portuguesas presentes em diversas línguas como o inglês, francês ou japonês. Por mais numerosas que sejam, estatisticamente são irrelevantes. Não obstante, há línguas que possuem palavras de origem portuguesa que se contam em mais de 90% do vocabulário total, sem, contudo, fazer com que essas línguas possam ser chamadas de português. Estas línguas, classificadas cientificamente como crioulos de base portuguesa, chegam a compor quase a totalidade de seu léxico com palavras de origem portuguesa. Um número tão grande não poderia se dar, obviamente, por um simples processo de empréstimo lexical. Nas línguas crioulas, as palavras portuguesas sofreram, como era de se esperar, todas as alterações fonológicas que toda palavra de qualquer língua sofre ao entrar em outra.

Cada língua tem a sua fonologia própria, seus fonemas, seu padrão de estrutura silábica, etc. Uma palavra inglesa como *sport*, monossilábica, existe tal como é, graças à estrutura das sílabas em inglês, estrutura essa que permite que uma sílaba nessa língua inicie com uma consoante fricativa alveolar surda, [s], seguida de outra consoante e terminada com uma consoante oclusiva alveolar surda, [t]. Tal regra não existe em português. Quando a palavra *sport* entrou na língua portuguesa, tornou-se trissilábica, devido a restrições fonotáticas do português. Da mesma forma, as palavras portuguesas chegaram aos crioulos de base portuguesa modificando seus fonemas e estruturas silábicas.

O estudo dos processos de transformação fonológica das palavras portuguesas que entraram no crioulo sãotomense pode nos dizer muito sobre o português do Brasil. Há

---

<sup>6</sup> O número é simbólico, levando-se em conta os 450 mil vocábulos do dicionário Houaiss, que não possui verbos conjugados ou substantivos no feminino e/ou plural. Contando-se todas as variações, o vocabulário da língua portuguesa sobe para a casa do milhão.

correntes acadêmicas, como Mendonça (1973) e Castro (2006), que defendem a teoria de que línguas africanas contribuíram para a formação do português do Brasil no campo da fonologia, morfologia e sintaxe, e não apenas no campo lexical.

Estudar as influências de línguas africanas como o quicongo (banto) e bini (edoide) nos crioulos do Golfo da Guiné pode nos ajudar a compreender os processos de transformação do português do Brasil, que esteve, por mais de três séculos, em contato constante com duas línguas bantas (quimbundo e umbundo) e uma ioruboide (iorubá), dentre outras.

No Golfo da Guiné falam-se quatro crioulos de base portuguesa: sãotomense, principense, angolar e anobonês. O sãotomense é uma língua classificada como crioulo de base portuguesa (com cerca de 150 mil falantes) e falada como língua nacional na República de São Tomé e Príncipe, país africano composto por duas ilhas principais, e tendo o português como língua oficial. O angolar é falado principalmente no sul da ilha de São Tomé, a principal ilha do país, onde se localiza a capital São Tomé, e conta com cerca de cinco mil falantes. O principense é falado principalmente na outra grande ilha da República de São Tomé e Príncipe, a ilha de Príncipe, e possui cerca de quatro mil falantes. O anobonês é falado na República da Guiné Equatorial, tanto na ilha de Ano Bom (atual Pagalu) onde surgiu, quanto na ilha de Fernando Pó (atual Bioko), onde se localiza a capital Malabo. Guiné Equatorial é uma ex-colônia espanhola, descoberta por portugueses, e que atualmente possui três línguas oficiais: o espanhol, o francês e o português, esta última oficial desde julho de 2007 (Asopgelibre, 2007). Sendo que o país não conta com um contingente significativo de lusófonos, a escolha do português como língua oficial deve-se mais a questões políticas e econômicas, como o desejo do país de ingressar na Comunidade de Países de Língua Portuguesa, do que a questões sociais ou históricas, conquanto a ilha de Ano Bom tenha sido colonizada por portugueses, e o crioulo anobonês, ainda falado, é um dos crioulos de base portuguesa e conta com mais de 90% de seu léxico de origem portuguesa.

Segundo Mane (2007), os quatro crioulos têm uma origem comum: são, mais precisamente, dialetos derivados do sãotomense. Mais do que essa tese controversa, o que nos levou a utilizar os outros três crioulos, além do sãotomense, foi o fato de contarem com palavras comuns em grande fatia de seu léxico.

O português que será analisado é a vertente que foi utilizada pelos portugueses entre os séculos XV e XVI, ou seja, o português europeu seiscentista. Usaremos como *corpora* os dados das obras de Coutinho (1976), Ferraz (1979), Silva (1991) e Teyssier

(2004). No século XVI, os portugueses mantiveram um império marítimo pela costa africana, Brasil e Ásia (China, Índia, Indonésia, Timor Leste). A língua portuguesa entrou em contato com línguas africanas, asiáticas e indígenas brasileiras, tomando por empréstimo palavras africanas (moleque, carimbo, samba), asiáticas (bambu, chá, jangada) e indígenas (pipoca, guaraná, peteca) e emprestando vocábulos para as línguas africanas, asiáticas e indígenas. Ao longo desse processo de difusão da língua portuguesa pelo mundo, surgiram novas línguas. Essas línguas, chamadas crioulos, receberam, durante sua formação, contribuições portuguesas no léxico e contribuições das línguas autóctones na fonologia, na morfologia, na sintaxe e na semântica.

### 1.3. Metodologia

Como *corpora* deste trabalho utilizei:

- a. A lista-Swadesh<sup>7</sup> levantada por Graham & Graham em 2004, detalhada na bibliografia<sup>8</sup>.
- b. A lista de vocábulos são-tomenses levantada por Mane entre 2003 e 2007 para sua tese de doutorado (Mane, 2007).
- c. A lista levantada por Ferraz em 1979.
- d. A lista levantada por Valkoff em 1966.
- e. A lista levantada por Negreiros em 1895 apud Ferraz (1979).

O trabalho será apresentado da seguinte maneira:

Na seção 1.4, apresentarei a geografia e história de São Tomé e Príncipe, como as línguas são faladas (língua materna, segunda língua ou língua oficial), e as políticas do idioma no país.

Na seção 2, **O Contato linguístico**, apresentarei a parte teórica sobre o crioulo são-tomense, começando com a definição de pidgins e crioulos; sua classificação entre as demais línguas vivas, mortas, naturais e artificiais; hipóteses correntes para a gênese dos pidgins e crioulos.

---

<sup>7</sup> Lista-Swadesh é uma lista de vocabulário básico (pronomes, numerais, partes do corpo, etc.) com palavras que dificilmente entrariam em outras línguas por empréstimo. Em glotocronologia serve para uma data aproximada da separação entre dois idiomas. Em dialetologia serve para indicar as diferenças fonéticas entre dialetos de uma mesma língua.

<sup>8</sup> Esta lista será usada como fonte principal deste trabalho, e para corrigir possíveis erros, será confrontada constantemente com outras listas já levantadas por outros pesquisadores, como as apresentadas nos itens de “b” a “e”.

Na seção 3, **Os substratos africanos dos crioulos do Golfo da Guiné**, farei uma breve apresentação das duas principais línguas que serviram de substrato para a formação do crioulo, isto é, as línguas maternas dos primeiros falantes do sãotomense, quando aprenderam o léxico português. É essencial conhecer as línguas de substrato para compreender o funcionamento fonológico e morfológico do sãotomense.

Na seção 4, **O superstrato português dos crioulos do Golfo da Guiné**, apresentarei uma visão sobre o português falado no século XVI, época da constituição dos crioulos, a fonologia do português, morfologia e demais aspectos para compreender o papel da língua portuguesa e a sua contribuição para a constituição do sãotomense, não apenas no léxico, mas também em outros campos da divisão clássica da língua.

Na seção 5, **Fonética e fonologia do sãotomense**, apresentarei os fones e fonemas do sãotomense, tanto consonantais quanto vocálicos, com base na lista-swadesh em comparação com Ferraz (1979). Por fim farei uma breve explanação da estrutura das sílabas em sãotomense e suas restrições fonotáticas. Confrontando com a fonologia das línguas africanas de substrato, apresentada na seção 3, e com a língua portuguesa, apresentada na seção 4.

Na seção 6, **Metaplasmos**, analisarei cada uma das adaptações fonéticas que as palavras portuguesas sofreram durante o processo de empréstimos para o sãotomense, tentando observar o motivo que levaram as palavras portuguesas a tal alteração.

Na seção 7, **Outras contribuições do substrato**, apresentarei em algumas páginas, a morfologia e a sintaxe do sãotomense, para uma visão mais aprofundada da língua como um todo, e não apenas na sua fonologia, demonstrando outras contribuições do substrato não apenas na fonologia.

Na seção 8, **Outras contribuições do superstrato**, analisarei o papel de contribuição da língua portuguesa na constituição do sãotomense, não apenas no léxico, mas também na fonologia e morfologia.

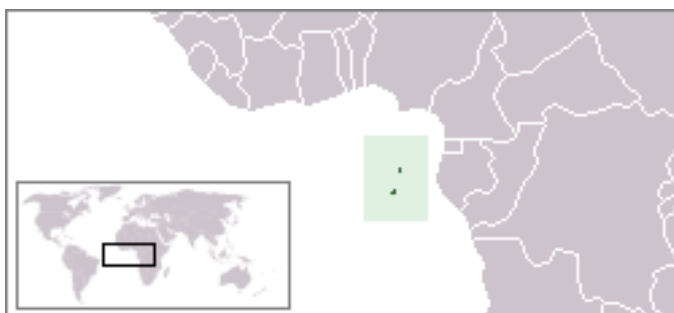
Na seção 9, **Contribuições conjuntas e elementos autênticos**, analisarei características presentes no sãotomense de forma que é impossível rastrear se a contribuição foi portuguesa ou africana, já que ambas apresentam as mesmas características. Apresentarei, ainda, características do sãotomense que não se relacionam nem com o substrato nem com o superstrato.

O trabalho será encerrado com a seção 10: **Considerações finais** e com a seção 11: **Referências bibliográficas**.

## 1.4. Geografia de São Tomé e Príncipe

### 1.4.1. Geografia física

A República de São Tomé e Príncipe se localiza, latitudinalmente entre a linha do Equador e o paralelo 1°42' acima dessa linha e, longitudinalmente, 7° a oeste do meridiano de Greenwich. A República de São Tomé e Príncipe é um arquipélago formado por duas grandes ilhas que dão o nome ao país e cinco ilhas menores: Pedras, Tinhosa Pequena, Tinhosa Grande, Rolas e Carço; além de dezenas de ilhotas, somando o total de 964 km<sup>2</sup> (cerca de dois terços do tamanho da cidade de São Paulo, que possui quase 1.523 km<sup>2</sup>).



**Mapa 1:** Golfo da Guiné, na África, com a localização das ilhas no retângulo ao centro. Mapa retirado de <[http://www.arikah.net/enciclopedia-portuguese/S%C3%A3o\\_Tom%C3%A9\\_e\\_Pr%C3%ADncipe](http://www.arikah.net/enciclopedia-portuguese/S%C3%A3o_Tom%C3%A9_e_Pr%C3%ADncipe)>.

Tanto as ilhas de São Tomé e Príncipe quanto as duas ilhas pertencentes à Guiné Equatorial: Pagalu (antiga Annobón) e Bioko (antiga Fernando Pó) estão alinhadas por uma falha geológica de origem vulcânica, a Linha de Camarões (Henriques, 2000: 24).



**Mapa 2:** As quatro ilhas do Golfo da Guiné alinhadas pela falha geológica da Linha de Camarões.<sup>9</sup>

Pode ser observada no mapa acima a existência de uma cadeia montanhosa entre Nigéria e Camarões. Esse acidente geográfico que se inicia no continente continua até o Oceano Atlântico, na mesma linha, estendendo-se, no fundo do mar. Entre o lado sudeste e noroeste da cadeia montanhosa estão duas placas tectônicas que se chocaram, gerando vulcões e montanhas, no continente; e ilhas, no mar. As montanhas no continente são o Maciço Adamawa, e as ilhas no mar, do nordeste ao sudoeste são: Bioko, onde se localiza a capital guineequatoriana, Malabo; a ilha de Príncipe; as minúsculas Tinhosa Pequena e Tinhosa Grande, invisíveis no mapa; a ilha de São Tomé, onde se localiza a capital São Tomé; a minúscula ilha das Rolas, invisível no mapa; e por último a ilha de Pagalu (Annobón), pertencente à Guiné Equatorial.

#### 1.4.2. Geografia humana

A população total do país atualmente está em cerca da 200 mil habitantes (estimativa para 2007 da Cia Factbook), divididos em diversas etnias, incluindo fangs e portugueses.

Há cinco principais grupos étnicos em São Tomé e Príncipe. Os angolares se concentram ao sul da Ilha de São Tomé; e os mestiços de Príncipe, na ilha de mesmo nome. Os fangs são uma etnia africana de língua banta muito numerosa e conhecida no

<sup>9</sup> Mapa disponível em <[http://upload.wikimedia.org/wikipedia/commons/thumb/a/a6/Gulf\\_of\\_Guinea\\_%28English%29.jpg/800px-Gulf\\_of\\_Guinea\\_%28English%29.jpg](http://upload.wikimedia.org/wikipedia/commons/thumb/a/a6/Gulf_of_Guinea_%28English%29.jpg/800px-Gulf_of_Guinea_%28English%29.jpg)>.



Golfo da Guiné. São 1,7 milhão na África e habitam, além de São Tomé e Príncipe, o Congo, Guiné Equatorial, Gabão e Camarões. Os fangs presentes no Golfo da Guiné são todos emigrantes da África Continental que se mudaram para o arquipélago lusófono em busca de empregos.

A etnia europeia mais numerosa é a dos portugueses, os brancos eurafricanos, que permanecem no arquipélago desde a colonização. São os falantes nativos da língua portuguesa, falada como língua materna por uma minoria, mas por 95% da população como segunda língua. Além do português, o são-tomense também funciona como língua veicular em todo o arquipélago (Smith, 2007 e Mane, 2007: 45).

**Tabela 1:** Os grupos étnicos de São Tomé e Príncipe.

<b>Etnia</b>	<b>Língua materna</b>	<b>Afiliação linguística</b>	<b>População</b>	<b>%</b>
Angolares	Angolar	Crioulo português	5.000	3,7 %
Fangs	Fang	Banto	12.900	9,5 %
Portugueses	Português	Neolatina / românica	2.580	1,9 %
Mestiços de Príncipe	Principense (moncó)	Crioulo português	4.000	2,9 %
Mestiços de São Tomé	Sãotomense (forro)	Crioulo português	110.000	81,7 %
<b>Total</b>			<b>134.480<sup>10</sup></b>	

Fonte: Smith (2007). Dados de 2001.

#### 1.4.3. Geografia linguística

O país atualmente é uma República Presidencialista, possui uma constituição promulgada em 1975, ano da Independência, e atualizada em 1990. Há uma língua oficial *de facto* — o português — e pelo menos quatro línguas nacionais (sãotomense, principense, angolar, fang). Não há, na constituição são-tomense, qualquer menção em relação à língua oficial, contudo, o português é usado na Justiça, no Executivo, nas Redações Oficiais e na Educação Básica. São chamadas línguas oficiais *de facto* aquelas usadas como língua oficial, embora não haja na constituição do país qualquer menção sobre língua com tal *status*. O mesmo ocorre, por exemplo, nos Estados Unidos, cuja constituição não prevê qualquer língua oficial para o trabalho do governo, mas um pacto

<sup>10</sup> Os dados são de 2001. A estimativa da Cia para 2007 é de 200 mil habitantes, como mencionado.

silencioso propõe o uso do inglês<sup>11</sup>. Em outros países, uma língua oficial é escolhida e mencionada na constituição, como é o caso do Brasil, país que possui quase duzentas línguas nacionais, mas tem apenas o português, língua materna de 88% da população, como língua oficial mencionada na Constituição de 1988 (art. 13)<sup>12</sup>. Tal escolha deve-se não somente ao passado histórico brasileiro como ex-colônia portuguesa, mas também devido ao fato de o português ser a língua materna de 88% da população.

Em São Tomé e Príncipe, a escolha da língua portuguesa se deveu a motivos semelhantes. Segundo dados de Sterzi (apud Smith, 2007), em sua Análise dos Dados do Recenseamento de 1981, 95% da população de São Tomé e Príncipe afirmou falar o português como segunda língua, contra 85,4 % que falava o sãotomense. Embora o multilinguismo seja comum entre africanos, e o sãotomense seja falado como segunda língua por grupos étnicos diversos como os principenses e angolares (Mane, 2007: 45), o sãotomense não foi escolhido como língua oficial. A necessidade de evitar confrontos étnicos motivados por ciúmes de grupos minoritários que não tivessem suas línguas escolhidas como oficiais, o fato de o português ser uma língua neutra e não ser língua materna de nenhuma etnia africana, levaram São Tomé e Príncipe à “escolha” do português como língua oficial *de facto*.



**Mapa 3:** Distritos de São Tomé e Príncipe: 1- Água Grande, 2- Cantagalo, 3- Caué, 4- Lembá, 5- Lobata, 6- Mé-Zóchi, 7- Pagué ou Príncipe.

<sup>11</sup> Alguns estados americanos possuem uma ou mais línguas oficiais, como a Flórida (inglês e espanhol), Louisiana (inglês e francês) e o Havaí (inglês e havaiano). Não há, contudo, na Constituição Federal, qualquer menção sobre língua oficial para o trabalho do Governo dos Estados Unidos e sua Justiça.

<sup>12</sup> Embora a língua oficial do Brasil seja o português, os estados e municípios podem escolher línguas cooficiais para o trabalho dos seus governos. É o caso do município de São Gabriel da Cachoeira, no Amazonas, com quatro línguas oficiais: Português, Nheengatu, Tukano e Baniwa (lei 145/2002, aprovada em 22/11/2002).

São Tomé e Príncipe é dividido, para efeitos administrativos, em sete distritos, sendo seis na maior ilha, a de São Tomé, e o sétimo distrito abarcando a totalidade da ilha de Príncipe. A capital, São Tomé, se localiza no primeiro distrito, Água Grande, o menor distrito do país, e onde reside mais da metade da população de São Tomé e Príncipe. Os três outros distritos mais populosos são justamente da área vizinha à capital, Mé-Zóchi, Lobata e Cantagalo.

**Tabela 2:** Porcentagem de falantes das três principais línguas.<sup>13</sup>

	<b>Distrito</b>	<b>Português</b>	<b>Sãotomense (forro)</b>	<b>Principense (moncó)</b>	<b>Outras línguas</b>
1	Água Grande	97,0 %	94,8 %	1,8 %	7,6 %
2	Cantagalo	97,3 %	78,4 %	1,3 %	22,1 %
3	Caué	89,3 %	78,8 %	0,5 %	17,9 %
4	Lembá	86,6 %	71,3 %	0,9 %	24,0 %
5	Lobata	94,4 %	78,5 %	0,5 %	21,8 %
6	Mé-Zochi	94,7 %	97,7 %	0,7 %	7,1 %
7	Pagué	97,7 %	71,7 %	24,6 %	37,2 %
	Zona Rural	94,0 %	80,5 %	2,7 %	17,5 %
	<b>Total</b>	<b>95,0 %</b>	<b>85,4 %</b>	<b>2,4 %</b>	<b>14,0 %</b>

Fonte: Smith (2007). Dados de 1981. Não inclui falantes abaixo de 10 anos de idade.

A difusão da língua portuguesa começa na capital. Os distritos mais próximos são os com as mais elevadas taxas de falantes de português, enquanto os distritos mais afastados, como Lembá e Caué, apresentam as menores taxas. A maior porcentagem de falantes de sãotomense se localiza em Mé-Zochi, próximo à capital, e em Água Grande, e a menor em Pagué (Ilha de Príncipe) e ao Sul da Ilha de São Tomé. Já os falantes de principense estão, em sua maioria, na Ilha de Príncipe, e em menor quantidade em Caué. Os distritos de Lembá, ao sul, e Lobata, ao Norte, apresentam as maiores taxas de falantes de outras línguas, sendo, provavelmente, o angolar em Lembá e o fang em Lobata.

### 1.5. História de São Tomé e Príncipe

Não se conhecem com rigor as datas da descoberta das ilhas de São Tomé e Príncipe. Tradicionalmente se admite que São Tomé tenha sido encontrada em 21 de

<sup>13</sup> As porcentagens indicam a proporção na população total de pessoas que saibam falar cada uma das línguas, não necessariamente como língua materna. Em Água Grande, por exemplo, 97% da população afirmou falar português. Como 94,8% da população fala sãotomense, a grande maioria é bilíngue ou trilingue.

dezembro de 1471, dia de São Tomé; Príncipe (antigamente chamada de Santo Antão) foi descoberta em 17 de janeiro de 1472, dia de Santo Antão; e a ilha de Annobón em 1 de janeiro de 1472, dia de Ano Novo (Henriques, 2000: 23). Essas suposições são feitas porque a prática portuguesa era batizar as novas terras encontradas com o nome do santo do dia. Uma exceção é a cidade do Rio de Janeiro, originalmente cidade de São Sebastião do Rio de Janeiro, descoberta em 1 de janeiro de 1502, dia de Ano Bom, de Santa Maria, São Justino, Santo Odilon e São Vicente, mas batizada em “homenagem” ao rei D. Sebastião.<sup>14</sup>

A ilha de Bioko era habitada por africanos, e houve certa resistência da população autóctone à colonização portuguesa (Henriques, 2000: 24). As outras três ilhas eram desabitadas, o que facilitou a colonização.

Em 1485 saiu de Portugal um pequeno contingente de colonizadores portugueses para se estabelecerem em São Tomé. Oito anos depois, mais um contingente de colonos, desta vez bem maior, se radicou na ilha (Ferraz, 1979: 9). Os europeus colonizadores eram, em sua maioria, degredados. Em 1501, foi introduzida em São Tomé a monocultura de cana-de-açúcar, sendo usada para isso mão-de-obra escrava oriunda da África continental (Pereira, 2007), esses escravos tinham como línguas maternas, possivelmente o quicongo ou o bini (Ferraz, 1979). Muito provavelmente, a flutuação da população era enorme, por ser São Tomé e Príncipe um entreposto de escravos. Por isso, dezenas de outras línguas podem ter contribuído na formação do crioulo são-tomense, embora seja difícil ou talvez até impossível rastrear todas essas línguas.

Esse contingente de escravos não era passivo e, tal como no Brasil com os quilombos, também houve revoltas em São Tomé e Príncipe. A mais famosa foi a de Amador, herói nacional são-tomense, com uma história curiosamente muito semelhante à do brasileiro Zumbi dos Palmares (1655-1695). Amador era um ex-escravo de um capitão do mato. Assim como Zumbi, também conhecia estratégia militar e técnicas de

---

<sup>14</sup> Em algumas versões, o batismo se deu em 1 de março de 1565, dia da fundação, o que explica a não correspondência entre 1 de março e São Sebastião (20 de janeiro). Para mais informações veja <[http://www.colorfotos.com.br/rio\\_de/princip.htm](http://www.colorfotos.com.br/rio_de/princip.htm)>. Segundo a versão mais aceita, como a mencionada pelo historiador Chico Alencar (<http://www.chicoalencar.com.br/chico2004/chamadas/2007/pronuc01032007.htm>), o batismo da cidade se deu após o término da batalha de expulsão dos franceses, em 20 de janeiro de 1567, dia de São Sebastião, que, segundo a lenda religiosa, apareceu lutando ao lado dos portugueses e índios tamoios. O verbete da Enciclopédia Britânica (Rio de Janeiro, 2008) parece tentar desfazer a contradição das versões: “When the foundations of the future town were laid in 1565, it was named Cidade de São Sebastião do Rio de Janeiro (“City of Saint Sebastian of Rio de Janeiro”), **for both** St. Sebastian and Dom Sebastian, king of Portugal”. (Grifo meu). A versão da Enciclopédia Britânica não resolve o problema, já que a batalha aconteceu em 1567, e o batismo, segundo essa fonte, se deu em 1565.

guerra. Autodeclarou-se rei de São Tomé em 1595, organizando um exército de africanos para conquistar a ilha de São Tomé, conseguiu libertar dois terços do território, inclusive a administração colonial localizada na capital. Assim como Zumbi, também foi traído por companheiros do exército, capturado pelos portugueses e executado (Pereira, 2007).

As revoltas, contudo, continuaram, levando à decadência do sistema escravagista, o que, associado à perseguição inglesa sobre a escravidão, levou à abolição da escravatura em 1876. Outros africanos chegam a São Tomé, mas desta vez trabalhadores livres, após a abolição.

No período que vai do fim do séc. XVI ao final do séc. XVIII a economia agrícola mostra-se estagnada, porém, com a introdução das culturas de cacau e café, que prosseguem até a independência, a agricultura recebe nova injeção de vigor. Mais mão-de-obra é requerida, porém, a escravidão já não se constitui como método válido e vantajoso para os colonos. Ocorre a abolição, imediatamente seguida pela introdução do contrato do trabalho, o qual trouxe a São Tomé enormes quantidades de negros angolanos, cabo-verdianos e moçambicanos.

(Pereira, 2007)

A independência de São Tomé e Príncipe só veio a ocorrer 99 anos após a abolição da escravatura, em 1975, ano da Revolução dos Cravos em Portugal.

Herança dessa colonização exploradora, São Tomé e Príncipe, atualmente, conta com uma economia majoritariamente agrícola, que produz cacau, coco, pimenta, café, banana, mamão, feijão e palma.

Para Ferraz (1979: 9) a colonização de Príncipe e Annobón se deu no século XVI com população oriunda de São Tomé. Devido a isso, o crioulo sãotomense arcaico (falado na ilha de São Tomé no século XVI), dividiu-se em quatro crioulos, devido à separação geográfica, mas também devido ao poder do substrato, que embora seja o mesmo nos quatro crioulos, é mais marcadamente banto em sãotomense e angolar, e mais marcadamente edoide no principense.

## 2. O CONTATO LINGUÍSTICO

### 2.1. O contato interlinguístico

Durante a história da humanidade, não foram raras as situações em que ocorreram contatos entre povos de línguas mutuamente ininteligíveis. Nesses contatos, tentava-se, emergencialmente, a comunicação a todo custo, por meio de gestos. A própria história do Brasil se inicia com a chegada de portugueses que, em contato com os índios pataxó (grupo macro-jê) na região onde atualmente se localiza o município de Porto Seguro – BA, enfrentaram uma situação potencialmente pidginizante, por não conhecerem uns a língua dos outros. Pero Vaz de Caminha assim descreve a situação:

Vio huũ deles huũas contas e rosairo brancas. Açenou que lhas desem e folgou muito com elas e lancouas ao pescoço e depois tirouas e enbrulhouas no braço e acenaua peraa trra e entã peraaas contas esperao colar do capitam como que dariam ouro por aquilo. Isto tomauamonos asy polo desejarmos mas se ele queria dizer que leuaria as contas e mais ocolar. Jsto nom queryamonos emtender porque lho nõ aviamos de dar edespo

(Caminha, 2002: 39).

A situação descrita na carta de Pero Vaz de Caminha conta um episódio desta situação potencialmente pidginizante, uma situação em que um índio e um português se comunicavam por meio de gestos. O índio acenava para o colar do navegante português, posteriormente à terra, dando a entender que pagaria ouro por ela. Caminha é bem honesto quando diz que entenderam isso porque assim o desejavam, mas se o índio queria dizer que levaria o colar, isso não queriam entender.

Assim como essa comunicação de condições precárias, muitas foram as situações na história do Império Português em que africanos de línguas mutuamente ininteligíveis entre si foram forçados a se comunicar com os portugueses e entre si sob risco até mesmo de inanição. A colonização das outrora desabitadas ilhas do arquipélago de São Tomé e Príncipe, para citar um exemplo conveniente, se deu dessa forma. Tanto São Tomé e Príncipe quanto Cabo Verde serviam de entrepostos de escravos. Africanos de etnias distintas e línguas ininteligíveis entre si eram agrupados nas ilhas, às vezes, por alguns anos, para posteriormente serem levados ao Brasil ou alhures. Numa situação como essa, a comunicação se tornou extremamente necessária. Os africanos tentaram aprender a língua do dominador, mas dadas as condições precárias dessa aprendizagem, principalmente no processo de transmissão da língua do povo dominador, esse aprendizado não poderia se dar de forma efetiva.

## 2.2. O processo de aquisição da linguagem

### 2.2.1. O processo monitorado de aquisição de uma língua estrangeira

No caso do aprendizado de uma segunda língua, sendo esse aprendizado monitorado (de forma organizada, com professores ou autodidaticamente por meio de material adequado) a ordem dos elementos (léxico, semântica, sintaxe, morfologia, fonologia e ortografia) pode variar, de acordo com a metodologia utilizada. Raramente os métodos de aprendizado de línguas estrangeiras seguem a ordem de aprendizado da língua materna, já que, em quase todos os casos, a ortografia da língua estrangeira é ensinada ao aluno já desde a primeira aula, junto com algumas frases já formadas, numa inversão da ordem natural de aprendizado. A fonologia, como pode ser facilmente observada, fica quase sempre prejudicada, pois o aprendiz toma sua língua materna como parâmetro para retirar os fonemas que utilizará na nova língua. Isso ocorre, por exemplo, quando um brasileiro ao aprender inglês, pronuncia [ˈkɛtʃi] para dizer “cat”, em inglês [ˈkæɪt].

No processo inicial de aprendizagem de uma outra língua, independentemente da metodologia aplicada e da ordem ensinada, é comum a utilização, por parte do aprendiz, de uma linguagem semelhante a um pidgin, em que o falante usa a própria língua como estrutura de base e, colando o léxico, fala a nova língua apenas substituindo as palavras. A esta linguagem se dá o nome de interlíngua. Corder (1978) demonstra que, apesar de pidgins e crioulos serem considerados línguas, o mesmo não pode se dizer das interlínguas por duas razões: a natureza dinâmica das interlínguas e o fato de que são linguagens relacionadas a indivíduos e não a comunidades (Corder, 1978 apud Lopes, 2001: 115). Para Bickerton (1975), essa interlíngua, chamada de *continuum* de aprendizagem linguística, como, por exemplo, o “anglo-espanhol” (espanhol falado por anglófonos em processo de aprendizagem da língua estrangeira) deve ter o mesmo direito ao título de sistema que as línguas como o inglês ou espanhol (Bickerton, 1975 apud Lopes, 2001: 118).

### 2.2.2. O processo não-monitorado de aquisição de uma língua estrangeira

Quando o aprendizado da língua estrangeira se dá de forma não-monitorada, ou seja, emergencialmente, de forma precária, sem um professor para ensinar, ou melhor, com “professores” informais transmitindo a língua alvo de forma precária, surge a situação perfeita para a formação de um pidgin.

A aprendizagem da primeira língua e da língua estrangeira se diferenciam, basicamente, porque no segundo caso o aprendiz já é falante fluente de uma língua, a sua língua materna. O adulto já possui sua língua materna gravada em seu cérebro, e a aprendizagem se dará por meio de relações e analogias da língua alvo com sua língua materna. Quando o aprendiz adulto é simplesmente obrigado a aprender uma nova língua por conta própria, numa situação pidginizante, tomando uma atitude ativa, e não uma atitude passiva, como numa sala de aula em que a língua lhe é apresentada, o falante precisa descobrir como funciona a nova língua, e a ordem de aprendizado parece seguir a da aquisição natural da primeira língua: 1. léxico, 2. semântica, 3. sintaxe, 4. morfologia. Sobre o aprendizado do léxico antes da sintaxe e da morfologia, Couto (1999) afirma:

O importante a reter é que a língua sempre começa pelo léxico (...). Na aprendizagem de L2 e no surgimento dos pidgins, bem como no dos crioulos, os aprendizes encadeiam várias palavras sem sintaxe, cuja relação umas com as outras só pode ser dada, quando é dada, pelo contexto da situação. Portanto, é como se se pegassem peças de diversas máquinas diferentes, e os aprendizes de L2 (ou os formadores do pidgin) tivessem que descobrir algo em comum entre essas peças a fim de formar uma nova máquina, com base no conhecimento que já têm de outras máquinas. Geralmente conseguem.

(Couto, 1999: seção 8.3).

Num aprendizado emergencial e em condições precárias, tendo-se a necessidade de aprender a língua do dominador, a ordem de aprendizado parece não mudar, o primeiro elemento que o dominado tenta aprender é o vocabulário. Palavra por palavra, o aprendiz vai memorizando como, na língua do dominador, se nomeiam os seres e as coisas. Após o léxico, aprendem-se a morfologia e a sintaxe, não necessariamente nessa ordem. Já a fonologia vem junto ao vocabulário, mas se a transmissão da língua alvo for precária, fonemas inexistentes na língua do aprendiz não serão aprendidos, e as palavras serão adaptadas, ocorrendo uma intromissão da fonologia da língua materna. Isso ocorre como no exemplo dado do falante brasileiro que pronuncia a palavra inglesa “cat” [kæt]



como [ˈkɛtʃi]. Se nunca for corrigido, a pronúncia se cristalizará, fazendo parte dessa interlíngua que aprendeu.

Essa ordem na aprendizagem da língua: léxico, morfologia e sintaxe, pode ser observada pela observação da gramática dos crioulos, como o sãotomense, em que o léxico é evidentemente português, a morfologia possui muitos elementos do português e do quicongo, e a sintaxe é mais próxima do quicongo do que do português (ver item 9.5). O que parece acontecer é que o aprendizado não se completou. Para Bickerton (apud Couto, 1999) um pidgin (que posteriormente se transformará em crioulo) é uma língua que consta de uma lista de palavras precariamente compartilhadas, usadas no modo pragmático, sem nenhuma sintaxe. Quando o aprendizado se completa em todos os níveis: fonológico, morfológico, sintático, semântico e lexical, salvo os problemas fonéticos do sotaque estrangeiro, o falante é capaz de falar fluentemente a língua alvo de sua aprendizagem (Couto, 1999). Em outras condições, por motivos diversos, esse processo não se completa. Essa língua num processo intermediário com léxico e sem sintaxe se internaliza, surge uma sintaxe e o pidgin se torna a língua materna de muitos, torna-se, enfim, um crioulo. Sobre a formação de uma linguagem emergencial numa situação pidginizante, como a dos portugueses em Porto Seguro em 1500, por exemplo, e sobre a semelhança entre a pidginização e aquisição da língua estrangeira, Couto (1999) diz:

Começando pela aquisição de L2 [língua alvo], trata-se de um processo em que o falante já domina uma L1 (sua língua materna). Vimos que, no estágio TIC [tentativa individual de comunicação], o processo se inicia apenas por palavras da língua alvo, sem sintaxe. Na melhor das hipóteses, essas palavras são usadas com **resquícios da gramática de L1**, sobretudo nos primeiros momentos. Quando se atinge o estágio de EIC [estratégias individuais de comunicação], e sobretudo quando começam a se fixar algumas ECC [estratégias coletivas de comunicação], o falante começa a aprender traços da gramática alvo também, podendo chegar a um de dois resultados. O primeiro, e mais comum em situações não monitoradas, é ele **adquirir L2 com a transferência de muitos traços fonéticos e gramaticais de sua L1**. O segundo, mais comum em situações de aprendizagem monitorada, consiste em apropriar-se de **L2 quase como falante nativo**. A formação dos pidgins, como vimos, se aproxima muito desse processo. Os seus formadores/usuários também têm sua L1 própria. A diferença fundamental consiste em que cada um deles pode ter uma L1 diferente.

(Couto, 1999)<sup>15</sup>.

---

<sup>15</sup> Grifo nosso. Os termos entre colchetes não constam do original.

### 2.3. Pidgin

Não há consenso entre os estudiosos sobre uma definição de pidgin. Couto (1994: 24) menciona que Mühlhäusler alinha seis definições, as de Bloomfield, Adler, Jespersen, Hall, Baesler e Reinecke, apresentando falhas em todas elas e dando em seguida sua própria definição, que Couto não menciona.

Bollée (1977 apud Couto, 1994: 28) caracteriza, e não define, os pidgins. Para definição, ficaremos com a de que pidgins são meios de comunicação (línguas) que surgem do contato entre povos distintos e obedece aos cinco critérios sociolinguísticos de Bollée.

Para Bollée (1977 apud Couto, 1994: 28) há dois tipos de critérios para definir um pidgin, os sociolinguísticos e os linguísticos. Segundo os critérios sociolinguísticos de Bollée, o pidgin:

1. É uma língua surgida do contato de dois ou mais povos de línguas mutuamente ininteligíveis. Frequentemente um dos povos é superior socioeconômica e politicamente.

2. Não é língua materna de ninguém (língua de contato).

3. É um meio precário de intercompreensão.

4. É um modo de comunicação pragmático, ainda não há uma gramática comunitariamente aceita.

5. Não apresenta nenhum sentimento de amor e fidelidade (sic) por parte de seus usuários. Assim que podem, abandonam-no.

### 2.4. Crioulo

Para Hall (1966), “um crioulo surge quando um pidgin se torna a língua nativa de uma comunidade de fala”.

Segundo esta definição, crioulos nada mais são, portanto, do que pidgins internalizados. Um dos problemas da definição de Hall é que, a princípio, só dá conta dos crioulos derivados de pidgins instáveis que passaram pelo estágio de um pidgin estável. Como é perfeitamente possível que um pidgin instável transforme-se num crioulo queimando etapas de pidginização e crioulição, como aconteceu com o crioulo do Havaí (Couto, 1994: 25) cruzaremos os dados de Hall com os de Mühlhäusler.

Couto cria um diagrama em triângulo, reproduzindo o ciclo-vital dos pidgins e crioulos de Hall. O choque entre povos de línguas ininteligíveis entre si leva à criação de um pidgin instável. Instável porque

o que se dá são estratégias individuais para comunicação interlinguística, (...) Não há nenhuma sintaxe, cada enunciado é proferido tendo por base, quando a tem, estruturas da língua original do falante (...) a comunicação se dá inteiramente pelo modo pragmático, ou seja, ela só se viabiliza devido ao contexto da situação. Não há estruturas subjacentes, tudo é improvisado no momento

(Couto, 1994: 25-26).

Esse pidgin instável, também chamado de “jargão” por Mühlhäusler (1986), pode pidginizar-se a ponto de tornar-se um pidgin estável. Tornando-se a língua materna de novas gerações, essas, ao adquirir a linguagem, dariam complexidade gramatical ao pidgin, regularizariam-no, e o mais importante, uniformizariam-no, transformando o pidgin num crioulo, Mühlhäusler (1986) dá a esse processo o nome de crioulição.

O processo poderia continuar, se a língua do dominador continuar presente na sociedade, como acontece hoje nos países africanos de língua oficial portuguesa, principalmente São Tomé e Príncipe, e o crioulo tenderia a se aproximar cada vez mais da língua do dominador, descrioulizando-se. A isso se dá o nome de descrioulização.

Nem sempre isso acontece. O pidgin estável pode desaparecer, devido ao aprendizado completo da língua do dominador, ou por condições sociais diversas, como a falta de necessidade de uso. A língua descrioulizada pode ainda fazer um processo de volta às origens (à língua crioula), o que se chama de recrioulição. Isso ocorre, geralmente, quando a sociedade crioula pretende reafirmar sua identidade étnica, tal como ocorreu com os jamaicanos que emigraram para a Inglaterra (Couto, 1994: 20).

Em resumo, para Mühlhäusler (1986, apud Arends et alii, 1995: 6; Couto, 1994: 25), um jargão (pidgin instável) pode transformar-se num pidgin estável, posteriormente num pidgin expandido, para finalmente se tornar um crioulo, caso ocorrido com o Tok Pisin, crioulo da Papua Nova Guiné. Pode ainda um pidgin instável chegar a um crioulo passando apenas pelo estágio de pidgin estabilizado, como o crioulo de base inglesa do estreito de Torres, ou ainda, um pidgin instável tornar-se um crioulo sem passar por essas etapas, como aconteceu com o crioulo de base inglesa do Havai.

Nesse sentido, seguiremos ao longo deste trabalho com a definição de crioulo como língua derivada de um pidgin instável, que, além das características já

mencionadas, também é uma língua criada num curto espaço de tempo, e que não guarda qualquer relação genética com outras línguas, conforme será visto na seção 3.6.

## 2.5. Substrato e superstrato

Como o empréstimo de terminologia é comum entre as ciências, a crioulistica tomou emprestados da linguística histórica os termos substrato e superstrato, que existiam para nomear, respectivamente, as línguas existentes na Europa antes da expansão romana, e o latim, língua oficial do Império.

Segundo uma controvertida teoria linguística hoje praticamente abandonada, o latim, ao se estabelecer numa província como a Península Ibérica, teria tido uma fase inicial de falar crioulo ao ser adotado pelas populações nativas

(Camara Jr., 1968: 176).

A teoria a qual Mattoso Camara se refere tratava o surgimento da língua portuguesa na Península Ibérica da mesma forma que a das línguas crioulas, vindas do choque de duas ou mais línguas, sendo o latim a língua do dominador, e as línguas autóctones dos povos pré-romanos de Portugal como as línguas do dominado. Nessa teoria, surgiu a terminologia de superstrato e substrato que foi tomada por empréstimo (Quint, 2007b: 4) pela crioulistica, parte da ciência linguística que estuda os crioulos, e a pidginística, que estuda os pidgins.

Para Camara Jr. (1968: 364) o substrato é o:

Nome que se dá à língua de um povo que é abandonada e esquecida em proveito de outra que a ele se impõe, em regra como consequência de conquista política. **O substrato persiste no léxico da nova língua**, que se enriquece com um resíduo de palavras, especialmente topônimos, e pode ainda aí introduzir traços morfológicos e fonéticos, estabelecendo-se assim uma modalidade *sui generis* de empréstimo linguístico. Assim, no léxico português, temos um pequeno substrato pré-romano em topônimos como *Coimbra*, *Lima* e em alguns nomes comuns, com *arroio*, *baía*, *cama*, *lousa*

(Camara Jr. 1968: 364).

Para definir substrato, Camara Jr. se utiliza da filologia românica, citando ainda as poucas palavras portuguesas que, etimologicamente, derivam de palavras das línguas extintas e desconhecidas que eram faladas na Península Ibérica, antes da chegada dos exércitos romanos e da língua latina. De fato, o substrato, também nos crioulos, deixa-se transparecer no léxico. Em são-tomense, por exemplo, cerca de 7% do léxico é de origem africana (Ferraz, 1979: 114-115). A quantidade reduzida de palavras se explica:

a língua do dominador é a principal doadora de vocábulos, por isso, na crioulistica, é chamada também de língua lexificadora.

Para definir superstrato, Camara Jr. diz que é o:

Nome que se dá à língua de um povo conquistador, quando ele a abandona para adotar a língua do povo vencido. O superstrato persiste no léxico da língua adotada, que se enriquece com termos referentes a traços específicos da cultura do povo conquistador. As línguas germânicas no território da România, com invasões bárbaras, constituíram superstratos do latim. Daí provém o apreciável acervo de germanismo nas línguas românicas

(Camara Jr. 1968: 368).

A partir de definições como essa, a crioulistica aproveitou apenas o fato de ser a língua do povo conquistador. Para a crioulistica, mais do que a língua do povo militar, política e economicamente mais forte, a língua de superstrato é aquela que cede a maior quantidade de palavras para o léxico da nova língua. No caso do são-tomense, o português, que cedeu cerca de 93% do léxico, é considerado a língua de superstrato.

Em toda situação crioulizante, há pelo menos uma língua de superstrato e uma ou mais línguas de substrato. Como os europeus foram os povos militar e tecnologicamente mais fortes durante a Idade Moderna, é natural que sejam as línguas europeias as línguas de superstrato na maior parte das línguas crioulas. Apesar disso, existem línguas crioulas com superstratos árabe e quicongo, dentre outras.

## 2.6. Os crioulos e a evolução das línguas

A língua portuguesa não pode ser considerada um crioulo de base latina principalmente porque os crioulos são línguas que se criam num curto espaço de tempo. O processo de criação da Língua Portuguesa foi lento e gradual: há 5.000 anos a.C., na região onde hoje é a Rússia, falava-se o proto-indo-europeu (Muysken & Smith, 1995: 4). O povo falante desse idioma teve sua língua gradual e lentamente modificada por processos naturais, a deriva linguística, levando a variações geográficas diversas depois que o povo indo-europeu se espalhou pela Eurásia. A variante geográfica falada por aqueles que se estabeleceram na Península Itálica evoluiu separadamente da variante dos outros grupos, até um nível em que a língua daqueles que se fixaram na Península Itálica já poderia ser chamada de proto-itálico. Várias línguas derivaram do proto-itálico. O latim, uma dessas línguas, ganhou prestígio e poder e, em sua expansão, marginalizou outras línguas da península. O latim, da mesma forma, transformou-se, ao

longo dos anos, no latim vulgar, para enfim, no século XII, transformar-se no português arcaico, na região ocidental da Península Ibérica. Obviamente essa linha evolutiva não foi reta, mas algo semelhante à linha da evolução biológica de Darwin: um tronco com múltiplos galhos e, desses galhos, outros ainda menores, de forma que línguas distantes como o português e o russo possam ser aparentadas entre si, tomando-se como base um ancestral comum mais distante que um ancestral comum ao português e o francês, por exemplo.

Esses nomes dados às línguas, como proto-indo-europeu, latim clássico, latim vulgar, português arcaico e português moderno são, na verdade, simbólicos e de natureza puramente didática. São nomes mais precisamente dados a diferentes estágios de uma mesma língua, que, num *continuum*, modifica-se gradualmente, transformando-se em línguas aparentadas, como português, francês, italiano, romeno, catalão, romanche, galego e espanhol apenas devido às variações do ancestral comum.

Apenas didaticamente a data do surgimento da língua portuguesa, 1196, é estabelecida. Na verdade, nunca houve rupturas ou saltos, a língua portuguesa se desenvolveu lenta e gradualmente.

O mesmo não se pode dizer dos crioulos. Esses são idiomas nos quais há uma visível ruptura no processo de evolução das línguas. Entre o português moderno e o proto-indo-europeu há uma linha contínua de evolução linguística mais ou menos rastreável. Nos crioulos, essa linha evolutiva é rompida no seu nascimento. Crioulos de base portuguesa não são línguas da família neolatina, muito menos do tronco indo-europeu (cf. item 3.7), tampouco pertencem a qualquer tronco de línguas africanas. São agrupados num tronco próprio, mais por analogias entre si que do por relações genéticas.

A palavra *crioulo*, em sua etimologia, vem do verbo criar, e significa língua que se cria<sup>16</sup> (Couto, 1994). Crioulos são línguas que nascem, são criadas em situações específicas e não possuem qualquer relação de continuidade evolutiva com as línguas de superstrato ou substrato. De fato, possuem o léxico em sua maioria do superstrato e o traços morfológicos e sintáticos predominantemente do substrato, com influências do superstrato, mas que podem ainda ser totalmente originais. Ainda assim, não derivam dessas línguas de substrato ou superstrato. Isso se deve ao fato de que a evolução não

---

<sup>16</sup> Há diversas hipóteses etimológicas para as palavras “pidgin” e “crioulo”, ficaremos com essa, que pode ser considerada tão improvável quanto todas as outras.

foi natural. Foi induzida, artificialmente, por um processo alheio às regras da evolução linguística.

Sobre a evolução histórica das línguas não poder ser usada para crioulos, Muysken & Smith (1995: 4) afirmam:

This kind of statement we can definitely not make when talking about creole languages. These exhibit an abrupt break in the course of their historical development. So we cannot say that Sranan (the major English-lexifier creole of Surinam) derives in any gradual fashion from Early Modern English...<sup>17</sup>

## 2.7. Os crioulos e a taxonomia linguística

A taxonomia linguística, parte da ciência linguística que cataloga e separa por afinidade todas as cerca de seis mil línguas conhecidas, não apresenta, como na biologia e nas ciências exatas, tamanha uniformidade. Tradicionalmente, a taxonomia linguística divide as línguas em cinco graus hierárquicos, do maior para o menor: tronco > família > grupo > subgrupo > língua, podendo, apesar disso, apresentar outros tantos graus intermediários e inominados quantos forem necessários, variando de pesquisador para pesquisador. Apesar das variações terminológicas entre os cientistas, existem algumas fontes como o site Ethnologue (Gordon, 2005) que conseguem manter um banco de dados satisfatoriamente uniformizado.

As mais de seis mil línguas conhecidas hoje estão catalogadas e incluídas em um dos 108 troncos existentes (Gordon, 2005f). Quando uma língua não se aparenta com nenhuma outra conhecida, por não possuir qualquer ligação genética, ainda sim um tronco lhe é reservado, o tronco das línguas isoladas. Dos 108 troncos linguísticos já catalogados, 101 são de línguas que apresentam alguma relação genética entre si, mas os outros sete troncos são de línguas sem relação aparente entre si, são eles:

- (1) tronco das línguas isoladas;
- tronco das línguas de sinais;
- tronco das línguas de sinais para surdos;<sup>18</sup>

---

<sup>17</sup> Este tipo de declaração (evolução das línguas) não se pode fazer quando falamos das línguas crioulas. Elas exibem uma quebra abrupta no curso de seu desenvolvimento histórico. Dessa forma, não podemos afirmar que o Sranan (o crioulo do Suriname com a maior quantidade de palavras de origem inglesa) deriva do Primitivo inglês moderno.

<sup>18</sup> A diferença entre os troncos das línguas de sinais (com apenas duas línguas) e o das línguas de sinais para surdos (com 114 línguas) está no fato de que as primeiras são usadas por não-surdos, apenas para comunicação em silêncio. São a língua *plain* dos indígenas americanos, para se comunicarem durante a

tronco das línguas artificiais ou construídas;<sup>19</sup>

tronco das línguas misturadas;

tronco dos pidgins;

tronco dos crioulos.

De todos os mais de cem troncos, apenas cinco já reúnem quase dois terços das línguas conhecidas: (1) Nigero-congolês, com 1514 línguas faladas na África; (2) Austronésio, com 1262 línguas faladas na Oceania e Extrema Ásia; (3) Trans-Newguine, com 552 línguas faladas na Papua Nova Guiné e Indonésia; (4) Indo-europeu, com 443 línguas faladas em todo o mundo; e (5) Sino-Tibetano, com 365 línguas faladas no Extremo Oriente.

A grande quantidade de línguas nos troncos nigero-congolês e austronésio se explicam pela diversidade na região geográfica na qual as línguas são faladas. Para que uma língua se divida em duas, há a necessidade de isolamento geográfico entre seus falantes, e a Oceania é farta nesta situação, com suas milhares de ilhas isoladas. O mesmo ocorre na África, um “arquipélago” continental onde muitos povos se isolam e perdem o contato entre si durante séculos, e quando se encontram, suas línguas já se distanciaram o bastante da língua comum que lhes deu origem.

No estudo da Linguística Africana, Greenberg menciona a existência, na África, de quatro troncos:

**Nigero-congolês**, com 1514<sup>20</sup> línguas catalogadas, dentre elas as do subgrupo banto. Dá-se o nome de bantuística à parte da Linguística Africana que estuda as mais de quinhentas línguas bantas, como o quimbundo de Angola, trazido ao Brasil; o quicongo, um dos substratos do são-tomense; e o swahili, a mais difundida língua africana, escrita com alfabeto latino e falada por mais de dez milhões de pessoas.

**Afro-asiático**, com 375 línguas catalogadas, dentre elas o árabe; o etíope (amárico), uma das poucas línguas não-europeias oficiais em Estados africanos; o hauçá

---

caça sem que os animais ouçam os caçadores, e a língua do monastério do Vaticano, usada por monges que queiram se comunicar em silêncio nos conventos.

<sup>19</sup> Mesmo agrupando línguas artificiais sem relação genética, algumas línguas construídas apresentam relação entre si, como o ido e o esperanto. A primeira foi criada com base na segunda. O mesmo pode ocorrer com o tronco dos crioulos, o fato de um tronco unir línguas sem relação, não impede que algumas delas possam, por acaso, ter uma origem comum.

<sup>20</sup> Segundo o Ethnologue 2008a são 1489 línguas, segundo Gordon (2005g) são 1514. A contradição das informações no banco de dados online se deve, provavelmente, à má organização e desatualização de algumas páginas. As diferenças numéricas ficam em até cinco línguas para mais ou para menos para cada tronco. Ficaremos com os dados mais numerosos em todos os troncos linguísticos.



ou haussá, que depois do swahili é a mais difundida língua africana, que conta com mais de 10 milhões de falantes, é escrita com alfabeto árabe; e o aramaico, que foi a língua administrativa e religiosa de diversos impérios da Antiguidade (Beyer, 1986).

**Nilo-saariano**, com 204 línguas catalogadas, dentre elas o kanuri, uma das 300 línguas que foram faladas no Brasil devido à importação de escravos do ciclo da Guiné, no século XVI (Petter, 2005: 199).

**Khoisan**, com 27 línguas catalogadas, dentre elas o !xóõ de Botswana e o !o!ung de Angola. Concentradas no sudoeste da África, possui bolsões em outras regiões mais ao norte. Há controvérsias entre os linguistas devido à necessidade de manter as línguas khoisan num tronco separado dos demais, e quanto aos critérios de reunião das línguas no mesmo tronco, sendo que, aparentemente, as línguas do tronco khoisan só se relacionam por um aspecto fonético/fonológico: a presença de cliques (sons característicos e, até o que se sabe atualmente, existentes apenas nas línguas africanas khoisan) (Heine, 2006). Foneticamente, segundo Graham e Graham (2004) há cliques também em sãotomense.

Apesar de serem apenas os quatro troncos os únicos mencionados tradicionalmente nos estudos de Linguística Africana, falam-se na África línguas indo-europeias, como o inglês, francês e português, que geralmente funcionam com línguas oficiais para uso na justiça, no governo, na divulgação científica e na educação. Faladas geralmente como segunda ou terceira língua pelos africanos.

Não podemos deixar ainda de mencionar o tronco dos crioulos, onde se encontram algumas línguas nascidas na África, de base portuguesa, inglesa ou francesa, para ficar apenas nas europeias. Esse tronco possui 81 línguas vivas catalogadas no mundo, sendo de 13 a 15<sup>21</sup> delas de base portuguesa (Gordon, 2005d), assim distribuídos geograficamente:

**No Golfo da Guiné (4):** angolar, principense e sãotomense (em São Tomé e Príncipe) e anobonês (Guiné Equatorial).

**Na Alta Guiné (2):** caboverdiano (Cabo Verde), crioulo (Guiné-Bissau)

**Na Ásia (6):** indo-português (Sri Lanca), kristang (Malásia), macanense (China), ternatenho (Indonésia), timor (Timor Leste), korlai (Índia).

---

<sup>21</sup> A incerteza numérica se deve ao fato de não existir consenso entre os linguistas sobre a existência ou não de uma base portuguesa para os crioulos papiamentu e saramaccan.

**Na América (0-3):** todas línguas de classificação controversa. Cafundó (Brasil)<sup>22</sup>, papiamentu (Antilhas Holandesas), saramaccan (Suriname).

### 2.7.1. A classificação genética

A classificação genética de uma língua pode ser dada num número indeterminado de degraus, em inglês há cinco, em português há dez. A classificação genética indica tanto a história da evolução da língua quanto a proximidade ou distância das línguas aparentadas entre si. No caso do português, vemos pela sua classificação genética (Indo-Europeu, Itálico, Românico, Ítalo-Occidental, Occidental, Galo-Ibérico, Ibero-Românico, Oeste-Ibérico, Português-Galego) que, de trás para frente, é uma língua do subgrupo Português-Galego, tendo como parentes mais próximas o galego e a fala de Xálima. Caso um lusófono se propusesse a aprender galego, poderia aprendê-lo mais facilmente, devido às muitas similaridades entre as duas línguas. O parentesco também nos faz pressupor que, quanto mais próximas, maior a probabilidade de que sejam inteligíveis entre si. Um texto escrito em galego, por exemplo, pode ser lido por um lusófono razoavelmente culto sem maiores problemas.<sup>23</sup>

No subgrupo superior Oeste-Ibérico, além do português, estão as línguas do oeste da Península Ibérica, como o espanhol e o mirandês, parentes um pouco mais distantes que o galego. No subgrupo Ibero-Românico, além das línguas já mencionadas, estão as línguas da Península Ibérica, como o catalão e o provençal. Assim, degrau por degrau, vemos o afastamento do português com outras línguas, por exemplo, em degraus: galego, espanhol, catalão, romanche, italiano, romeno, até chegar ao tronco indo-europeu, com línguas aparentadas muito distantemente, como o albanês, inglês, lituano, polonês e línguas da Índia. Línguas de outros troncos, como Nigero-congolês ou Afro-Asiático não são aparentadas com o português.

A classificação genética do saótomense, pelo Ethnologue (Gordon, 2005) é apenas: Creole, portuguese base. Contudo, outras classificações já foram propostas. O saótomense poderia ser classificado, por exemplo: Crioulo, base portuguesa, Golfo da

---

<sup>22</sup> O cafundó, falado em Sorocaba, SP, é mais precisamente um crioulo invertido, ou anticrioulo, já que o léxico é banto e não português. Dessa forma, de um modo mais radical, podemos dizer que não existem crioulos de base portuguesa na América. Isso reduz ainda o número de crioulos no mundo para oitenta.

<sup>23</sup> Exemplo de texto em galego: “Toda persoa ten os dereitos e liberdades proclamados nesta Declaración, sen distinción ningunha de raza, cor, sexo, idioma, relixión, opinión política ou de calquera outra índole, orixe nacional ou social, posición económica, nacemento ou calquera outra condición” (Declaración Universal dos Dereitos das Persoas, disponível em: <<http://www.unhchr.ch/udhr/lang/gln.htm>>).

Guiné, são-tomense. Já que no Golfo da Guiné são falados outros crioulos como o principense, o angolar e o anobonês, todos mais próximos do são-tomense do que o kristang da Malásia ou o papiamentu do Caribe.

## 2.8. Hipóteses para o surgimento dos pidgins e crioulos

O gramático normativo e crioulista Celso Cunha (1980), em seu artigo “O protocrioulo português e a sua universalidade nos séculos XVI, XVII e XVIII”, apresentado em Goa, Índia, em 1978, no International Seminar on Indo-Portuguese History, traça a história das hipóteses de surgimento dos crioulos, não apenas portugueses, mas também de outras bases europeias.

Mais recentemente, em 2007, no 3º encontro internacional do Grupo de Estudos de Línguas Africanas, na USP, o linguista francês Nicolas Quint (2007b) reapresentou, contribuindo com suas atualizações, essas diversas teorias sobre a origem das línguas crioulas já resumidas por Cunha (1980) e outros linguistas.

Grosso modo, as teorias sobre a origem das línguas crioulas podem ser divididas em quatro grupos: teoria monogenética, baby-talk, bioprograma e gíria dos marinheiros.

### 2.8.1. A teoria monogenética

Todos os crioulos de base europeia, independentemente do superstrato, possuem semelhanças na fonologia, morfologia, sintaxe e no léxico. Isso levou os primeiros pesquisadores a aventar a hipótese de que alguns crioulos surgiram de uma única língua, de base lexical portuguesa.

O início dessa hipótese veio com a publicação de Whinnom (1956 apud Cunha 1980), um estudo comparativo de três crioulos de base espanhola falado nas Filipinas: ermitanho, cavitenho e zamboanguenho, que, segundo sua teoria, provinham todos do ternatenho, língua franca malaio-portuguesa.

Mas a ousada hipótese de uma origem única para todos os crioulos do mundo veio a lume com Thompson (1959 apud Cunha, 1980), que afirmava que todos os crioulos descendiam de um protocrioulo português quinhentista, formado na região das feitorias. Para Naro (apud Cunha, 1980), esse protocrioulo teria se originado em Portugal. A teoria vai mais além, ao indicar que talvez essa protolíngua tivesse formado também os

crioulos de base inglesa, francesa e holandesa, já que existem palavras portuguesas entre eles.

A teoria monogenética parte do pressuposto de que todas as línguas crioulas tiveram uma origem comum. Para isso, os adeptos dessa teoria se atêm na busca das semelhanças existentes em todos os crioulos, como, por exemplo, a regularização do sistema verbal, em que não existem desinências de número ou pessoa para indicá-las, mas uma forma única de conjugação, diferenciando-se as pessoas com o uso obrigatório de um pronome sujeito: **m-sabi**, **bu sabi**, **e sabi** – “eu sei”, “tu sabes”, “ele sabe”, etc. como em caboverdiano (Quint, 2007b: 1). O primeiro ponto que justificaria essa teoria é o da redução do paradigma sintético, pois nas línguas crioulas não há concordância verbal/flexão verbal. O segundo ponto é a oposição entre os verbos de estado e os verbos de ação.

Até mesmo o léxico, traço de ligação dos crioulos com seus superstratos europeus e traço distintivo dos crioulos entre si, pelo menos considerando os crioulos de base europeia, apresenta características que, à primeira vista, sustentam essa teoria, como a presença de vocábulos portugueses em crioulos de base, até onde se sabe, não portuguesa.

**Tabela 3:** léxico krio de origem inglesa X léxico krio de origem portuguesa<sup>24</sup>.

Inglês	Krio anglicizado	Krio conservador	Português
know	<i>no</i>	<b>sabi</b>	conhecer (< saber)
basket	<i>baskit</i>	<b>blai</b>	cesta (< balaio)
baby	<i>bebi</i>	<b>pikin</b>	criança (< pequeno)

**Tabela 4:** léxico papiamentu de origem portuguesa X léxico papiamentu de origem espanhola<sup>25</sup>.

Português	Papiamento port	Papiamento esp	Espanhol
<i>deixar</i>	<b>desha</b>	<b>laga</b>	<i>largar</i>
<i>esquecer</i>	<b>skeze</b>	<b>lubidá</b>	<i>olvidar</i>
<i>costumar</i>	<b>kustumá</b>	<b>kustumbrá</b>	<i>acostumbrar</i>
<i>murchar</i>	<b>morcha</b>	<b>marchitá</b>	<i>marchitar</i>

A presença de vocábulos portugueses nos crioulos de base inglesa e espanhola pode indicar, contudo, um simples empréstimo lexical comum a qualquer língua do mundo. Portugueses estiveram presentes em Serra Leoa, área de fala krio, e no Caribe

<sup>24</sup> Krio é um crioulos de base inglesa falado em Serra Leoa, noroeste da África. Exemplos retirados de Quint (2007b).

<sup>25</sup> Papiamento é um crioulo de base controversa (supostamente espanhola) falado no Caribe holandês. Exemplos retirados de Quint (2007b).

holandês, área de fala papiamentu, tanto que o nome do país africano e da ilha caribenha de Curaçao têm origem portuguesa, respectivamente “de **serra** em forma de **leoa**” e “ilha do **coração**” (Otero, 2006: 177, 77). Se os portugueses estiveram presentes em tais locais, como indicado pelos topônimos, a ponto até mesmo de batizar os países com palavras de sua língua, natural seria que empréstimos lexicais tivessem ocorrido, tanto que ainda hoje existem centenas de vocábulos nas línguas japonesa, inglesa e francesa de origem portuguesa, sem que essas línguas descendam do português.

A hipótese monogenética, de natureza eurocêntrica, foi levantada quando se olhavam apenas para os crioulos de base europeia. Naturalmente, os crioulos com base em línguas não-europeias, como o árabe e o quicongo, enterraram definitivamente essa teoria.

### 2.8.2. Baby-talk

A teoria do baby-talk, por sua vez, defende que os falantes da língua de superstrato (lexificadora) simplificaram (regularizaram) voluntariamente e conscientemente a própria língua para facilitar a comunicação com aqueles que desconheciam sua língua (futuros falantes do crioulo). Tal como os adultos comumente fazem com bebês, ao deturpar a própria língua principalmente no nível fonológico, acreditando que os bebês o compreenderão melhor, também os portugueses assim fizeram com os africanos. Essa teoria explicaria muitas semelhanças dos crioulos com a linguagem infantil, no que diz respeito, por exemplo, à ausência de morfemas verbais e uso dos infinitivos como se fossem formas conjugadas. Já em relação às palatalizações, despalatalizações e lambdacismos, como no exemplo são-tomense: “relógio” > lo<sup>h</sup>lozu, ou à presença de ideofones: “muito quente” > kêtʃi-kêtʃi, a semelhança com a linguagem infantil é apenas uma coincidência. Palatalizações e despalatalizações são comuns em várias línguas. Esses fenômenos ocorreram no português e no francês em relação a latim, e o mesmo ainda ocorre no português do Brasil. Já o lambdacismo em são-tomense, é, provavelmente, influência do quicongo, língua africana que não possui fonemas vibrantes alveolar e consoantes velares.

### 2.8.3. Bioprograma

Possui pontos em comum com as ideias da gramática universal de Chomsky (Quint, 2007b). Sua grande vantagem é explicar desenvolvimentos semelhantes em crioulos muito diferentes no que diz respeito ao seu modo ou época de formação. Segundo essa teoria, todo crioulo tende a se “simplificar”, todos os elementos menos marcados tendem a desaparecer, o que explica a tendência às sílabas CV, o desaparecimento de fonemas como interdentais, etc.

Para os chomskyanos, os crioulos se formam espontaneamente, não por escravos nascidos no continente, mas pela população nascida nas ilhas, aproveitando as estruturas gramaticais inatas com as quais todo ser humano nasce.

Para Bickerton (1975), os crioulos são criados pelas crianças nas localidades onde só se ouve pidgin, sem estrutura suficiente de qualquer língua natural. As crianças usam, então, a capacidade inata da linguagem para transformar o input deficiente numa língua plena.

Essa teoria gerativista explica, por exemplo, como é possível que crioulos falados em regiões tão distantes entre si e com base lexical diferente tenham estruturas gramaticais semelhantes. São as capacidades linguísticas inatas que tornam os crioulos semelhantes.

### 2.8.4. Gíria de marinheiros

A teoria da gíria dos marinheiros parte dos dados de muitas palavras existentes em crioulos que remetem, etimologicamente, diretamente à gíria falada pelos marinheiros na época das grandes navegações, como a palavra “nobega” em kabuverdianu, que vem de “navega” e significar “dar-se bem com alguém” (Quint, 2007b).

Segundo essa teoria, os crioulos tiveram como superstrato o português falado por marinheiros analfabetos, que não dominavam o português padrão da época. Muitas simplificações gramaticais dos crioulos, segundo essa teoria, são heranças das simplificações que os próprios marinheiros faziam, involuntariamente, em suas linguagens.

### 2.8.5. Conclusão

As teorias apresentadas não são, necessariamente, contraditórias entre si, e talvez todas tragam uma base empírica, misturada com alguns pontos de difícil comprovação. Os crioulos de base portuguesa, assim como o português do Brasil, tiveram como base o português falado por marinheiros (cf. teoria da gíria de marinheiros) e camponeses saídos de Portugal. Falantes esses que cometiam desvios em relação à norma gramatical, principalmente em relação à concordância e ortoepia. Isso talvez explique a falta de concordância existente não só nos crioulos como também no português do Brasil, como tentaram demonstrar Scherre e Naro, que encontraram a origem da falta de concordância de gênero e número no português popular brasileiro e no português de Helvécia não em línguas africanas, mas em textos europeus em português arcaico (Scherre & Naro, 2001).

Além do uso de uma variante não-padrão por parte dos marinheiros, possivelmente, esses mesmos marinheiros analfabetos ainda tentaram simplificar a própria língua para facilitar a comunicação com africanos (cf. teoria do baby-talk). Soma-se a isso a evolução natural das línguas, pois um crioulo formado no século XVI, como qualquer língua existente, teria que sofrer as alterações fonológicas, morfológicas e lexicais, que qualquer língua sofreria ao longo de quatro séculos (cf. teoria do bioprograma), sobretudo não tendo, ao lado da língua oral, uma cultura letrada, com escrita, literatura e gramática normativa para desacelerar o processo de deriva linguística (cf. o problema da deriva, item 3.9).

### 2.9. O problema da deriva linguística

A colonização de São Tomé e Príncipe começou em 1485. Pelos cálculos de Ferraz (1979: 5 e 31), a formação do crioulo são-tomense se deu num período de 80 anos, aproximadamente entre 1490 e 1570.

Some mediaeval features in the Portuguese component of São Tomense, together with other linguistic and historical evidence (...) suggest that the Creole was formed within a matter of decades; allowing a period of approximately 80 years for the formation and consolidation of the language, this would make São Tomense about 400 years old.

(Ferraz, 1979: 5)

Numa sociedade em que as mulheres provavelmente tinham seus primeiros filhos antes dos vinte anos, em 80 anos poderiam ocorrer pelo menos quatro gerações: filhos, netos, bisnetos e trinotos de falantes de quicongo, bini e dezenas de outras línguas. Três gerações são mais do que suficiente para que uma língua se crioulixe.

Se esses dados estiverem corretos e o são-tomense tenha surgido há cerca de cinco séculos, estaremos lidando com uma língua que possui cerca de 500 anos de história, mas só foi devidamente documentada há menos de 30.

A língua portuguesa dispõe de uma gramática desde 1536, uma literatura vasta e valorizada, uma epopeia em decassílabos publicada em 1572, uma escrita eficiente em alfabeto latino e uma cultura letrada. Apesar de todas essas forças centrípetas, que desaceleram a deriva linguística, alterações naturais que todas as línguas sofrem na sua evolução (Bagno, 2006), ainda assim a língua portuguesa mudou em cinco séculos, fonética, fonológica, morfossintática e lexicalmente.

Ferraz (1979: 9) afirma que uma vez que um crioulo é estabelecido, ele não está mais tão suscetível às mudanças quanto as outras línguas (note-se o léxico português ainda presente no anobonês). Apesar de Ano Bom ter sido cedido à Espanha em 1778, o léxico não foi facilmente substituído por hispanholismos.



### 3. OS SUBSTRATOS AFRICANOS DOS CRIoulos DO GOLFO DA GUINÉ

Como o português é a língua lexificadora dos quatro crioulos do Golfo da Guiné, 93% do léxico do sãotomense é de origem portuguesa (Negreiros, 1895: 355-369 apud Ferraz, 1979: 8). As áreas em que o português e os crioulos diferem (seja no campo fonológico, morfológico, sintático ou semântico) devem-se em muitos casos às influências das línguas dos grupos bantoide ou edoide da costa oeste africana. Escravos que chegaram a São Tomé vieram da baía do Benim (área de fala edoide) e do reino do Congo (área de fala banta, atualmente pertencente à República Popular de Angola) (Ferraz, 1979: 12-14).

O bini (também conhecido como edo) e o quicongo são as línguas de maior influência nos crioulos do Golfo da Guiné, como o sãotomense (de substrato mais quicongo que bini) e o principense (de substrato mais bini que quicongo).

**Tabela 5:** Quadro comparativo: as línguas de São Tomé e Príncipe e seus substratos:

Nome da língua ao longo deste trabalho	Outros nomes correntes na literatura (1)	Região onde se fala	Número de falantes	Substrato (4)
Sãotomense	Sãotomense, lungwa san tome, san tome, lungwa di tela, forro.	Ilha de São Tomé	110.000 (2)	<b>Kongo, bini</b>
Principense	Lun'gwiye, iye, moncó, lungwa d'iyé	Ilha de Príncipe	4.000 (3)	<b>Bini, kongo</b>
Angolar	Lungwa ngola, ngola, lwanda	Sul da ilha de São Tomé	5.000 (4)	Kimbandu, ndigi, <b>kongo, bini</b>
Português		Todo o país (oficial)	2.580 (5)	

(1) Mane (2003).

(2) Segundo Gordon (2005b) são cerca de 70 mil falantes. Os dados de 110 mil são 2001, retirados de Smith (2007) cerca de 163 mil. Números aproximados. Inclui falantes de sãotomense como segunda língua.

(3) com seus 200 falantes (dados de 1999, segundo Gordon, 2005b) ou 4.000, cálculo sobre Smith (2007) o principense é hoje uma língua em perigo de extinção.

(4) Gordon (2005b).

(5) Segundo Gordon (2005b), há 2.580 falantes de português como língua materna, o mesmo número para Smith (2007). São contados apenas os falantes de português como língua materna.

### 3.1. Quicongo

**Quicongo** – classificação genética (Nígero-congolês, Atlântico-Congo, Volta-Congo, Benuê-Congo, Bantoide, Meridional, Strictu Bantu, Central, H, Kongo)<sup>26</sup>.

Também chamado de kikongo, kikoongo e kongo, esta língua africana do grupo banto, com três milhões de usuários, é falada pelos bacongo, nação africana originalmente habitante das florestas equatoriais do sudeste africano, hoje divididos em três países: Angola, onde residem quase 1,2 milhão de falantes de quicongo; República Democrática do Congo (antigo Zaire), onde residem 1 milhão de falantes; e República do Congo, onde estão 836 mil falantes. Há ainda alguns milhares de falantes de quicongo habitando o sul do Gabão.

Angola, onde foi fundado o primeiro reino bacongo, é de onde provavelmente vieram os falantes de quicongo que inicialmente povoaram São Tomé e Príncipe. O quicongo atual falado em Angola é uma língua que descende do quicongo arcaico que serviu de substrato para os crioulos de São Tomé e Príncipe no século XVI, como o sãotomense e o principense e também crioulos da América, como o palenquero. Foi ainda a base para o quituba, um crioulo banto falado como língua franca em parte da África Ocidental (Gordon, 2005e).

No século XVII foi também uma das línguas faladas no Brasil, como língua materna, a partir do segundo ciclo de importação de escravos (o ciclo do Congo e de Angola), que trouxe para o Brasil, além do quicongo, outras duas línguas bantas, o quimbundo e o umbundo (Bonvini & Petter, 1998: 73).

Formas cristalizadas do quicongo arcaico são usadas hoje em rituais religiosos não apenas no Brasil, com a vertente do candomblé conhecida como candomblé Angola, mas também na Jamaica e em Cuba.<sup>27</sup> Contudo, muitos candomblecistas brasileiros estudam e falam o quicongo moderno, numa tentativa de recuperar as raízes da religião

---

<sup>26</sup> Gordon (2005c). Para mais informações sobre o significado da classificação genética, ver item 3.7.1. No caso do quicongo e do edo, são línguas tão aparentadas entre si quanto o português e o romanche, já que quicongo e edo estão na mesma família (Benuê-congo) mas não na mesma subfamília (bantoide X edoide), assim como o português e o romanche estão na mesma família (Gallo-Iberian) mas não na mesma subfamília. O quicongo tem como parentes mais próximos o swahili (ambos do subgrupo Narrow Bantu, sendo o quicongo Southern e o swahili Central), mas como parente mais próximo ainda o quimbundo (ambos do subgrupo Narrow Bantu Southern).

<sup>27</sup> A informação me foi passada oralmente por um sacerdote de religião afro-brasileira.

conhecida como candomblé de Angola, e não mais o quicongo arcaico trazido ao Brasil no século XVII.

As primeiras descrições do quicongo foram feitas pelos padres italianos Bonaventura de Sardegna, em 1645, e Hiacinto Brusciotto de Vetralla, em 1659. Durante a colonização belga no Congo, antigo Zaire, muito se publicou sobre a língua quicongo, sobretudo na Bélgica, como as gramáticas didáticas para aprendizado autodidata de Dereu (1955) e Coene (1960). Posteriormente, linguistas como Lumwamu (1973) contribuíram com descrições científicas. A descrição feita pelos padres italianos está muito mais próxima do quicongo arcaico que serviu de substrato do sãoatomense e que era falado no século XV do que a descrição do quicongo moderno feita por linguistas como Dereu (1955), Coene (1960) e Lumwamu (1973). Não obstante, as primeiras descrições tradicionalmente pecam pela ausência de rigor científico e terminológico, algo justificável devido aos conhecimentos da época. Ficaremos com a descrição feita por Lumwamu (1973), confrontando-a com as demais descrições quando for necessário para preencher as possíveis lacunas. Apesar de poder ter havido alterações fonéticas ao longo da história da língua quicongo, uma descrição do quicongo moderno não será de todo estranha ao objetivo de compreender o quicongo arcaico, pois o tempo entre uma língua e outra são demasiadamente curtos para que tenha havido alterações substanciais. Além disso, algumas características presentes tanto no quicongo moderno quanto no sãoatomense e ausentes no português podem mostrar que determinados elementos da língua se mantiveram inalteráveis.

**Tabela 6:** inventário fonológico das consoantes do quicongo.

	labiais	labiodentais	dentais- alveolares	palatais	velares
Oclusivas	p b		t d		k
Nasais	m		n	ɲ	ŋ
Fricativas		f v	s z		ʃ
Laterais			l		
Africadas		pf [bv]	ts dz		
Pré-nasais	mp mb	mf mv	nt nd		nk ng

Traduzido do quadro de Lumwamu (1973: 43). Lumwamu descreve o /k/ como oclusiva, mas o coloca no quadro das fricativas, por considerá-lo contraparte de [ʃ], ou seja [χ]. Reorganizamos os fonemas /k/ e /ʃ/ na tabela. Talvez o que tenha motivado a confusão de Lumwamu seja a chave para compreender a fricativação de [k] para [χ] em anobonês: cavalo > χabalu (Barrena, 1973).

Pelo que se pode observar, não há nada que se assemelhe aos róticos do português europeu seiscentista: [r] ou [ʀ] ou do português brasileiro moderno: [χ], [h], [ɦ], [ʀ]. O mais próximo é o fonema [ɣ] que existe no português do Rio de Janeiro como alofone de [r] em posição de coda e antecedendo consoantes sonoras: *carga* [ˈkaɣgɐ] (Silva, 1999: 38). Em quicongo, contudo, tal consoante não pode ocupar a mesma posição, já que não existem sílabas pesadas (Coene, 1960: 1; Lumwamu, 1973: 25). A única saída para o sistema fonológico são-tomense foi utilizar a consoante alveolar com modo de articulação mais próximo, no caso, o [l], o que explica o lambdacismo existente em são-tomense, conforme será aprofundado na seção 8.2.4.

Lumwamu (1973) não menciona [ʃ] e [ʒ] como fonemas, embora sejam sons extremamente comuns em quicongo. A não menção no quadro de fonemas se deve à inexistência de oposição em pares mínimos. Em quicongo, existe um processo fonético de distribuição complementar entre as consoantes não-palatais /s/, /z/, /ts/, /dz/, /ns/, /nz/ e suas contrapartes palatais: [ʃ], [ʒ], [tʃ], [dʒ], [nʃ], [nʒ], respectivamente. As palatais só ocorrem antes de [i], e as não-palatais ocorrem antes de [a], [e], [o], [u], [w]. Fones palatais não ocorrem antes de vogais [a], [e], [o], [u], [w], e fonemas não-palatais jamais ocorrem antes de [i]. Nos dizeres de Lumwamu (1973: 33, 34): “dans le contexte /s + i/, en position initiale ou intervocalique /s/ a un effect acoustique intermédiaire entre /s/ et [ʃ] palatalisé. (...) dans le contexte /z + i/, en position initiale ou intervocalique, /z/ s’entend comme un [ʒ] palatalisé.” O mesmo é informado por Tim (2008), que cita como exemplos de consoantes não-palatais, em sua posição característica: **tobola** “abrir um buraco”, **tanu** “cinco”, **kesola** “ser cortado”, **kunezulu** “céu”, **nswetu** “nosso”, **zevo** “então”, **nselele** “cupim”; e de consoantes palatais, em sua posição característica: **zima** “esticar”, **tjina** “cortar”, **tjiba** “banana”, **nkofi** “leão”, **azimola** “esmola”, **lolonzi** “fazer faxina”. Quase todos os mesmos exemplos são encontrados no são-tomense, segundo Ferraz (1979: 52).

Esse sistema de distribuição complementar é semelhante ao existente em são-tomense, cujo léxico sofreu palatalizações e despalatalizações em sua formação, seguindo a regra que só permite palatais antes de [i], proibindo-os nas outras situações. Assim, em são-tomense, temos casos em que as consoantes [s] e [z] se palatalizaram antes de [i]: “cinco” > **ʃinku**, “música” > **ˈmuzika**, e se mantiveram antes de outras vogais: “cego” > **ˈsegu**, “azedo” > **ˈzedu** assim como casos em que [ʃ] e [ʒ] se

despalatalizaram antes das outras vogais: “bicho” > **'bisu**, ferrugem > **fa'luza** mas se mantiveram antes de [i].

A diferença entre o sistema de complementação quicongo e o sãotomense está no fato de que o sãotomense não proíbe as palatais antes de vogais que não [i], embora isso não negue a hipótese de que a palatalização em sãotomense seja influência do sistema fonético do quicongo, como defendia Ferraz (1979). Em sãotomense, as palatais, com raras exceções, só ocorrem antes de vogais que não [i] quando, na etimologia, puder ser rastreado um [i] que tenha sofrido síncope. Originalmente, a consoante não palatal se palatalizou, vindo a encontrar-se com uma vogal que não [i] após a queda desta:

(6) **Português > intermediário > sãotomense**

Palácio > \*pa'laʃju > pa'laʃu

Bacia > \*ba'ʃia > ba'ʃa

Gênio > \*'ʒjɛnu > 'ʒɛnu

(Ferraz, 1979: 23)

O fenômeno de adaptação fonética de palavras estrangeiras, que ocorreu durante a formação do léxico sãotomense, também ocorreu com o quicongo. Assim como em sãotomense houve despalatalizações antes de vogais que não [i], como no caso de “José” > zɔ'zɛ (Ferraz, 1979: 69), em quicongo também houve o mesmo fenômeno, como nos empréstimos franceses: “Joseph” > zózé:fo (Lumwamu, 1973: 34). O fenômeno das palatalizações e despalatalizações será aprofundado adiante, nas seções 6.2.2 e 6.2.3.

O sistema vocálico do quicongo pode ser resumido da forma a seguir, baseada em Lumwamu (1973):

**Tabela 7:** inventário fonológico das vogais do quicongo.

		arredondada posterior	média anterior	não-arredondada anterior
abertura	Mínima	/u/		/i/
	2 <sup>a</sup>	/o/		/e/
	3 <sup>a</sup>		/a/	

Não há vogais nasais (Ferraz, 1979: 49). Todas as sílabas em quicongo são leves (abertas), ou seja, sílabas CV (Lumwamu, 1973: 25). Coene já mencionava em relação às sílabas CV, inclusive, a questão das consoantes pré-nasais, que poderiam ser mal interpretadas pelos europeus, dando a entender que existiriam sílabas fechadas em consoante nasal: “toute syllabe se termine par une voyelle: Ex. Muntu (homme) ne se prononce pas mun-tu, mais mu-ntu.” (Coene, 1960: 1). As consoantes duplas como ‘mb’, ‘mf’, ‘ns’, ‘ts’ são sempre consideradas como um único fonema, e são indivisíveis.

O quicongo possui um sistema tonológico com dois tons: um tom baixo, representado a seguir com acento grave, e um tom alto, representado a seguir com acento agudo.

- (7) /bákà/ “muro”                      /bàkà/ “alcançar”  
/díá/ “comer”                      /dìà/ “este que”  
/fúlù/ “espuma”                      /fùlú/ “lugar”

Os tons são distintivos lexicalmente (Lumwamu, 1973: 26). Também existe distinção lexical entre as vogais longas e breves:

- (8) /bàkà/ “alcançar”                      /bà:ká/ “limpar uma palmeira”

Em quicongo não existem ditongos (Coene, 1960: 1), o que pode explicar o fenômeno das monotongações existentes em são-tomense: peito > <sup>h</sup>petu, conforme será aprofundado na seção 6.4.2.1.

Algumas características morfológicas podem ser resumidas, baseando-se em Coene (1960): o sistema morfológico é como o das demais línguas bantas, possui um sistema de classes morfossintáticas. Os adjetivos se localizam após os substantivos,

como em português e são-tomense: **bantu bambote** “os homens bons”, literalmente: “homens de bondade”. Os pronomes pessoais são: **mono** (1SG), **ngeye** (2SG), **yandi** (3SG), **beto** (1PL), **beno** (2PL), **bau** (3PL). Os pronomes adjetivos possessivos ou demonstrativos se localizam também após o substantivo, assim como em são-tomense. É a ordem inversa ao português: **muntu oyu** “homem este”.

### 3.2. Bini

**Bini** – classificação genética (Nigero-congolês, Atlântico-Congo, Volta-Congo, Benuê-Congo, Edoide, Norte-Central, Edo-Esan-Ora)<sup>28</sup>.

Mais conhecida como edo, mas também chamada de bin, benin, addo, oviedo, ovioba, essa língua com mais de um milhão de falantes e que outrora deve ter sido falada no Benin (atualmente só o é na Nigéria) serviu de base secundária para crioulos como o são-tomense e base principal para o principense.

Na década de 1970, Luiz Ivens Ferraz e Wilfred Günther descreveram dois dos crioulos de São Tomé e Príncipe, o são-tomense e o principense, respectivamente, e para tal citaram constantemente a língua bini, que naquela época se classificava como kwa, assim como o iorubá. Atualmente, tanto o bini quanto o iorubá são, depois dos estudos de Bennett & Sterk (1977), reclassificados geneticamente na família Benuê-congo, da qual também fazem parte as línguas do subgrupo banto.

Falado atualmente por um milhão de pessoas na Nigéria, e por uma minoria no sul da ilha de São Tomé (Eweka, 2008), o bini é uma língua histórica do reino do Benin. Segundo a crença corrente entre o povo edo, o bini é uma língua pré-histórica, falada na África desde 1.200 aC (Igbineweka, 2008b). O bini atualmente goza de grande prestígio na Nigéria, onde é usado na educação (desde a básica até a superior) e conta com programas de rádio e TV, além de um dicionário e uma tradução completa da Bíblia feita em 1996 (Gordon, 2005a). Descendente direta do bini seiscentista que serviu de base principal para o principense, também influenciou bastante o são-tomense e o angolar, além do anobonês, outro crioulo de base portuguesa falado no Golfo da Guiné. De estrutura silábica leve, com um grande número de sílabas CV, como a maioria das línguas africanas, exerceu forte influência fonológica nos crioulos de base portuguesa.

---

<sup>28</sup> Gordon (2005a).

Uma das descrições fonético-fonológicas do bini foi feita por Thomas (1910: 133), que listou 34 fonemas (ou fones) consonantais. Seu trabalho é uma das primeiras tentativas de descrição científica da fonologia bini. Thomas (1910) chega a mencionar sua impossibilidade de mapear satisfatoriamente a fonologia bini, devido à sua falta de tempo, além da grande variação dialetal de tribo para tribo.

Numa descrição fonológica para leigos, no Edo Language Dictionary Online (2008) menciona que o alfabeto oficial do edo possui 7 vogais, 17 consoantes simples e 8 consoantes duplas, o que, sem terminologia adequada, a torna inadequada para uma análise abrangente. Em outra descrição ortográfica e fonológica do bini, Hartell (1990) lista 25 fonemas consonantais. Apesar de não mencionar ponto ou modo de articulação, há o uso de símbolos fonéticos da Associação Fonética Internacional (IPA). As dificuldades de organização dos fonemas de Hartell (1990) numa tabela pode ser minorada com o auxílio dos dados de Thomas (1910), apesar da falta de rigor, e da descrição superficial, mas extremamente rigorosa de Ladefoged (1993). Por fim, os dados da fonologia de Hartell (1990) serão confrontados com as lições da língua bini de Igbineweka (2008a) e a descrição para leigos de Edo Language Dictionary Online (2008).

**Tabela 8:** sistema consonantal do bini.

	bilabial	labiodental	dental	alveolar	velar	glotal	labiovelar
oclusiva	p b		t d		k g		kp gb
nasal	m		n				
vibrante							
fricativa		f v	s z	ʃ ʒ	x ɣ	h	
semivogal	w				j		
aproximante				ɻ ɻ			
lateral				l			
pré-nasal	mw						
não-pulmônica		β					

Os fonemas na tabela acima são baseados no cruzamento dos dados de Thomas (1910), Hartell (1990), Ladefoged (1993) e Igbineweka (2008-a)

Thomas (1910: 133) lista 34 fonemas consonantais, destes, alguns não foram listados na tabela deste trabalho por falta de informação específica por parte do autor.



Trocamos de posição a oclusiva (explosive) aspirada [h], e excluímos um par de fonemas sem ponto de articulação que Thomas localiza entre as velares (gutturals) e palatais, grafando k/g com um circunflexo subscrito; e um par t/d sem ponto de articulação que Thomas coloca entre as dentais e labiodentais, registrando t/d simples (enquanto grafa t/d com um ponto subscrito para indicar as dentais). Thomas cita ainda dois pares de palatais, um de africadas e outro de fricativas. Os fonemas não são mencionados por nenhum dos outros autores, motivo pelo qual os excluímos. Thomas menciona a existência de dois pares de nasais dentais que grafa com ‘n’ e ‘l’ e diacríticos. Tais consoantes não encontram correspondentes no IPA. As consoantes bilabiais não-pulmônicas são chamadas por Thomas de ‘inspirates’, cita um par, mas a correspondente surda [p] não existe no IPA. A consoante inspirada [b], muito provavelmente era uma implosiva vozeante bilabial [ɓ], contudo, Ferraz (1979: 49) afirmava não existirem consoantes implosivas em quicongo ou bini. Hartell (1990) se propõe a representá-la com < vb >. Hartell cita o par de oclusivas labiovelares kp/gb, assim como Igbineweka (2008-a), mas Thomas não o faz, talvez por acreditar serem fonemas separados. As alveolares líquidas que Thomas (1910) e Hartell (1990) se confundem em mencionar são descritas por Ladefoged (1993), que menciona o contraste fonológico existente entre quatro alveolares líquidas: alveolar lateral sonora [l], alveolar aproximante sonora [ɭ], alveolar fricativa sonora [ɮ], e alveolar aproximante surda [ɮ̥]: [álázi] “macaco” X [áɭába] “borracha”; [aɮá] “lagarta” X [àɮà] “parte da cerimônia de sepultamento”.

O sistema vocálico possui sete vogais orais (Hartell, 1990: 224):

**Tabela 9:** sistema vocálico bini I.

Vogais orais			
	Anteriores	Central	Posteriores
Altas	i		u
Médias-altas	e		o
Médias-baixas	ɛ		ɔ
Baixas		a	

Todas as sete vogais orais do bini possuem contrapartes nasais (Ferraz, 1979: 49):

**Tabela 10:** sistema vocálico bini II.

<b>Vogais nasais</b>			
	Anteriores	Central	Posteriores
Altas	ĩ		ũ
Médias-altas	ẽ		õ
Médias-baixas		ẽ	õ
Baixas		ã	

O sistema oral é, coincidentemente, o mesmo sistema vocálico que o português apresentava no século XVI. As vogais abertas nasais não existem em sãotomense, mas aparecerão no crioulo em que a base bini foi mais forte do que o quicongo, o principense.

Quicongo e bini contribuíram como substrato com três crioulos do Golfo da Guiné. As influências de substrato foram, além do léxico minoritário já mencionado, também no campo na fonologia, morfologia e sintaxe.

#### 4. O SUPERSTRATO PORTUGUÊS DOS CRIoulos DO GOLFO DA GUINÉ

O português, língua materna de 210 milhões de pessoas, segunda língua de 30 milhões e língua oficial de nove estados soberanos, possui mais de oitocentos anos de história, sem contar o período de séculos nomeado pelos filólogos como fase pré-histórica da língua portuguesa.

É impossível definir com precisão e nomear as fases de evolução da língua portuguesa. Diversos autores já fizeram periodizações, nomeando as diversas fases da evolução da língua. Alguns autores como Vasconcelos (1959), Câmara Jr. (1975) e Coutinho (1976) dividiam a história da língua portuguesa em apenas duas fases, excluindo-se aqui o período pré-histórico, as duas fases são: Português Arcaico, que vai do surgimento dos primeiros documentos em português (século XII ou XIII) até meados do século XVI, e Português Moderno, do século XVI aos nossos dias. Outros dividem esse primeiro período em duas fases, a primeira terminando em 1350, fim da produção lírica portuguesa, fase do galego-português (Teyssier, 2004 e Cuesta, 1961) ou português antigo (Cintra, 1963), quando começa a nova fase, até o século XVI, fase do Português Pré-Clássico (cf. Galves, 2006 e Ribeiro *et alii*, 2007). Said Ali (1921) já mencionava, no início do século passado que:

Alterações linguísticas não dependem do calendário, nem do ano em que o século acaba ou começa. O que devemos entender por linguagem quinhentista, seiscentista, etc., é a maneira de falar dominante em grande parte da respectiva era, ou nela principalmente. Dizeres peculiares a qualquer das épocas continuam muitas vezes a ser usados por alguns escritores do período seguinte.

(Said Ali, 1921).

O que nos interessa no presente trabalho, em relação à língua portuguesa, a princípio, é descrever o funcionamento fonológico, morfológico, sintático e lexical de uma variante diacrônica usada pelos marinheiros portugueses entre os anos de 1470 e 1570, aproximadamente. Foi nesse primeiro século de colonização que, segundo Ferraz (1979: 31) surgiu o crioulo são-tomense, que poderíamos também chamar, em nova nomenclatura filológica, de são-tomense arcaico, já que não apenas não é o mesmo falado atualmente, como também serviu para formar outros crioulos como o angolano, o principense e o anobonês (Ferraz, 1979: 9 e Mane, 2007).

Ainda sobre as divisões conflitantes entre as fases da língua portuguesa, o que todos os autores concordam, sobretudo, é que o século XVI foi um divisor de águas na

evolução da língua portuguesa. De fato, do final do século XV até o final do século XVI, os portugueses se espalharam pelo mundo, com as grandes navegações. Tivemos, nessa época, o descobrimento de São Tomé e Príncipe (1470) e do Brasil (1500), e no século XVI a publicação da primeira gramática da língua portuguesa, de Oliveira (1536) e do poema épico *Os Lusíadas*, de Camões (1572). A língua falada pelos navegantes portugueses, nesta fase, possuía características tanto do português arcaico quanto do português moderno. Tanto o português do Brasil quanto os crioulos africanos conservam muitos traços medievais que desapareceram na Europa.

Havia, ainda, as variações diastráticas (a língua do navegante português analfabeto não era a mesma do nobre alfabetizado de Lisboa), e as variações diatópicas (a língua de norte de Portugal, de Lisboa e de Açores e suas variações regionais).

Muitas diferenças entre o português arcaico e o moderno podem ser assinaladas na fonética, na morfologia, na sintaxe e no léxico:

#### 4.1. Características do português arcaico

##### 4.1.1. Fonética

Havia perfeita distinção entre a realização fonética dos grafemas <s> e <ç>, <s> intervocálico e <z>, <ch> e <x>, como, por exemplo, em coser [ko<sup>h</sup>zer] e cozer [l<sup>h</sup>ko<sup>h</sup>dzer]. Ver tabela 12.

Pronúncia [õ] para o atual ditongo [ãw], nos substantivos que em latim terminavam em –one e –udine e nos verbos que terminavam em –unt. Sermone > **sermom** > sermão. Solitudine > **soidom** > solidão. Amarunt > **amarom** > amaram.

Existência de diversos hiatos que posteriormente se desfizeram por crase ou ditongação. Escaecer > **esqueecer** > esquecer. Sedere > **seer** > ser.

Nasalidade em vogais, por influência dos m ou n originais. Luna > **lũa** > lua.

O sufixo –vel tinha a forma –bil ou –vil. Terríbil, semelhávil (cf. Coutinho, 1976: 66).

#### 4.1.2. Morfologia

Os nomes terminados em *-nte*, *-or* e *ês* eram uniformes: a infante, mha senhor, língua português. Havia plural para substantivos que, atualmente, não fazem plural: *ourívezes*, *alférezes*, *arráezes*. Muitos substantivos tinham gênero diferente: a fim, a mar, a planeta, a cometa, o tribo, o coragem, o linguagem. (cf. francês: *la fin*, *la mer*, *la planète*, *la comète*, e espanhol: *el coraje*, *el lenguaje*).

O morfema verbal da segunda pessoa do plural era *-des*. (cf. os verbos monossilábicos no infinitivo, *vós vedes*, *vós ledes*, em que a forma permanece atualmente). O morfema do particípio passado dos verbos da segunda conjugação era *-udo*: *perdudo*, *conhoçudo*, *escondudo*. (cf. *Conteúdo*, *teúdo*, *manteúdo*, formas em que o morfema permaneceu). Dentre outras diferenças. (Coutinho, 1976: 66-67).

#### 4.1.3. Sintaxe

As diferenças, nesse âmbito, eram mais numerosas. Dentre outras, podemos citar o uso do pronome oblíquo do caso nominativo para indicar acusativo ou dativo, e vice-versa (Coutinho, 1976: 67), tal como ainda ocorre no português brasileiro, nas variantes populares: “E o senhor disse...que enforcariam ell” (cf. o português atual: e o senhor disse que o enforcariam [erudito] X e o senhor disse que enforcariam ele [popular]). “O coração pode mais ca mim” (cf. a forma atual: o coração pode mais que eu [corrente em todas as variantes do português] X o coração pode mais que mim [que não ocorre mais em nenhuma variante, atualmente]).

#### 4.2. Algumas características do português seiscentista

O português que era falado durante o século XV e XVI possuía algumas características medievais que ainda conviviam com outras características que se sobreporiam às anteriores. Entre o final do século XV e o final do século XVI muitas alterações ocorreram no campo da fonologia. Algumas dessas características podem nos ajudar a datar a formação do sãotomense, assim como compreender algumas alterações fonológicas que as palavras portuguesas (do século XVI) sofreram ao entrar no sãotomense.

#### 4.2.1. Fonologia

O português que serviu de base lexical para os quatro crioulos do Golfo da Guiné era falado por marinheiros e degredados nos séculos XV e XVI. As diferenças diatópicas, diafásicas e principalmente diacrônicas para o português moderno, explicam muitas das características peculiares nos crioulos estudados. Assim como o português do Brasil se manteve mais conservador em diversos aspectos fonológicos e sintáticos do que o português europeu, também os crioulos do Golfo da Guiné mantiveram características arcaizantes, que podem ser observadas não apenas no léxico, como também na fonologia.

No léxico, palavras como “feder” do português arcaico, deu /fe<sup>1</sup>de/ em sãotomense e no português do Brasil, mas desapareceu em Portugal, sendo substituída por “cheirar mal”, ou ainda “jogar”, do português arcaico, deu /zu<sup>1</sup>ga/ em sãotomense, e [ʒo<sup>1</sup>ga] no português do Brasil, mas foi substituída por “atirar”, “lançar” no português europeu moderno. Há ainda palavras do português arcaico que se mantiveram em crioulos, mas desapareceram tanto no Brasil quanto em Portugal, como [pĩ<sup>1</sup>sa], com o significado de empurrar (Ferraz, 1979: 103), e origem no português arcaico “pinchar” (cf. o espanhol ‘pinchar’ – furar, picar, estimular).

Fonologicamente, os crioulos também se mantiveram mais conservadores. Uma palavra como “chuva” pronunciava-se [tʃuba] em português seiscentista, e manteve a mesma pronúncia em caboverdiano (Graham & Graham, 2004: 9).

##### 4.2.1.1. O sistema consonantal

Em 1536, justamente a época do surgimento dos crioulos do Golfo da Guiné, Fernão de Oliveira publicou a primeira gramática da língua portuguesa. Para Coseriu (2000), as descrições articulatórias do sistema consonântico feitas por Oliveira (1536) são completas e surpreendentemente exatas, e, salvo a terminologia, aceitáveis ainda hoje. Oliveira indicava também as consoantes com o mesmo traço articulatório mas com sonoridade diferentes: “... trata-se das letras: c-g; p-b; t-d; ss-s; ç-z; f-v; x-j, ou seja, dos fonemas: k/g; p/b; t/d; ś/ź; s/z; f/v; š/ž” (Coseriu, 2000: 40).

#### 4.2.1.2. As sibilantes

Os símbolos *ś/ź* que Coseriu utilizou se referiam aos fonemas ápico-alveolares [ʃ] e [ʒ] e os símbolos *š/ž* aos fonemas palatais [ʃ] e [ʒ].

Os grafemas <s, ss> e <ç> que são atualmente pronunciados como [s], os grafemas <s> intervocálico e <z>, atualmente pronunciados como [z] e os grafemas <ch> e <x> pronunciados como [ʃ] possuíam, no português arcaico, valores distintos e não se confundiam como hoje (Coutinho, 1976: 66 e Teyssier, 2004: 60).

Na idade média, havia os fonemas /ts/ *cen*, e /dz/ *cozer*. Os elementos oclusivos iniciais /t/ e /d/ começaram a se perder, e por volta de 1500, as oclusivas /ts/ e /dz/ já haviam se transformado em /s/ e /z/, assim /tsen/ > /sen/ e /codzer/ > /cozer/.

Os grafemas <ss> e <s>, para Oliveira (1536) pronunciavam-se da seguinte forma: “o s singelo, diz Quintiliano, é letra mimosa, e, quando a pronunciamos, alevantamos a ponta da língua para o céu da boca e o espírito assobia pelas ilhargas da língua. O ss dobrado pronuncia-se como o outro, pregando mais a língua no céu da boca.” (Oliveira, 1536 com grafia atualizada in: Buescu, 1975). Ou seja, eram consoantes apicais<sup>29</sup>, pronunciadas com a aproximação do ápice da língua com o palato.

**Tabela 11:** o sistema das sibilantes nos século XVI.

	Pré-dorsodentais <sup>30</sup>	Ápico-alveolares
Surdas	/s/ escrito <ç>, e <c> antes de <e> e <i>	/ś/ escrito <s> e <ss>
	Ex. paço	Ex. passo
sonoras	/z/ escrito <z>	/ź/ escrito <s>
	Ex. cozer	Ex. coser

Retirado de Teyssier (2004: 60).

<sup>29</sup> Consoante apical: aquela cuja articulação é realizada com o ápice (ponta da língua) próximo ou colado ao palato duro, alvéolos ou dentes incisivos superiores (p.ex.: o [r] "caipira" do port. do Brasil em porta; o /r/ do inglês em dry 'seco'; o/s/ do esp. [da Espanha] em casa etc.); consoante ápico-alveolar: consoante em que a ponta da língua toca ou se aproxima dos alvéolos dos dentes incisivos superiores (p.ex.:/t/ e/d/ do ing. em table, day); consoante ápico-dental: aquela que se articula com a ponta da língua próxima ou colada aos dentes incisivos superiores (p.ex.:/t/ e/d/ do esp.) (Martins, 2006).

<sup>30</sup> As duas pré-dorsodentais eram pronunciadas com a ponta da língua virada para baixo, e a parte anterior do seu dorso próxima dos dentes de cima. As duas ápico-alveolares eram pronunciadas com a ponta da língua próxima dos alvéolos. (Teyssier, 2004: 60).

Palavras como “paço” e “passo” continuavam pronunciando-se distintamente, esta com um fonema áptico-alveolar aquela com um pré-dorsodental. Palavras como “cozer” e “coser” também não se confundiam, sendo a primeira pronunciada com um fonema pré-dorsodental e a segunda com um áptico-alveolar. Meio século depois, os grafemas se confundiriam fonologicamente, transformando-se em homófonas palavras como “passo” e “paço”, “cozer” e “coser” (Teyssier, 2004: 60-61). No final do século XVI, depois que o crioulo são-tomense já estava formado, confusões começam a aparecer nos textos entre as pré-dorsodentais e áptico-alveolares que lhe correspondem. “Paço” e “passo” se tornaram homófonas com /s/, e “cozer” e “coser” com /z/, tal como é hoje.

**Tabela 12:** evolução da pronúncia das sibilantes em Portugal.

	Séc. XV	Séc. XVI	Séc. XVII
Paço	pa[ts]o	pa[s]o	pa[s]o
Passo	pa[s]o	pa[ʃ]o	pa[s]o
Coser	co[z]er	co[z]er	co[z]er
Cozer	co[dz]er	co[z]er	co[z]er

A partir de Teyssier (2004).

O fonema áptico-alveolar [ʃ] fica entre o alveolar e o alveopalatal. Para Teyssier (2004: 60): “um ouvido francês percebe nela um começo de ‘chiamento’, como se se tratasse de um fonema intermediário entre o [s] de *casser* e o [ʃ] de *chocher*.” O mesmo se pode dizer do áptico-alveolar [z], intermediário entre [z] e [ʒ]. Esse fonema áptico-alveolar intermediário entre [s] e [ʃ] é o mesmo existente em quicongo antes de [i] (cf. seção 4.1).

#### 4.2.1.3. Os fones [ʒ] e [ʃ]

Durante um período mais extenso, que foi até meados do século XVII, não havia confusão entre os grafemas <ch> e <x>. O primeiro pronunciava-se como atualmente é em espanhol [tʃ], e o segundo, como é atualmente hoje [ʃ] (Teyssier, 2004: 64-65). Isso, e não um processo de mudança fonética, explica a atual pronúncia de chuva, em caboverdiano [tʃuba], que manteve a pronúncia corrente no século XVI e XVII. Já em são-tomense, o fonema [tʃ] se transformou em [s] antes das vogais que não [i]: chuva >



**suba.** O sãotomense não conservou o fonema oclusivo, como o caboverdiano, apesar de ter se formado no século XV e a transformação de [tʃ] em [ʃ] ter se dado muito posteriormente, somente nos século XVII (Teyssier, 2004: 64).

O [s] e o [z] em posição de coda são pronunciados em boa parte do Brasil como alveolares, sendo a realização fonética surda ou sonora de acordo com a situação da sílaba seguinte: “faz tudo” [fas ˈtudu], “faz direito” [faz diˈrejtʊ]. As mesmas consoantes são palatalizadas na pronúncia normal em Portugal e algumas regiões do Brasil, como Rio de Janeiro, Santos, São Vicente e Belém: “faz tudo” [faʃ ˈtudu], “faz direito” [faʃ diˈrejtʊ]. Em sãotomense, as sibilantes também se palatalizam na posição de coda: fresco > **ˈfleʃku**, máscara > **ˈmlaʃka** (Ferraz, 1979: 34, 47). Ferraz pressupõe que a palatalização já fosse comum nos séculos XV e XVI, mas Teyssier (2004: 67) menciona que, se ela fosse comum nessa época, compreende-se mal por que no português do Brasil, na sua forma mais comum (salvo o Rio de Janeiro, Santos, etc.), não exista a palatalização do [s] e do [z] em coda. Para Teyssier, os [s] e [z] em coda seriam, inicialmente, sibilantes (como ocorre em espanhol e no português do Brasil em sua forma mais comum), e em época mais tardia (entre o século XVI e XVIII) teria se palatalizado. Isso explica por que o português do Brasil não palataliza, mas não explica por que em sãotomense o [s] e o [z] em coda sejam palatalizados.

#### 4.2.1.4. A flutuação entre [b] e [v]

No capítulo em que descreve as transformações sofridas pelo português, de meados do século XIV até os nossos dias, Teyssier (2004: 59) menciona a atual falta de distinção entre os grafemas <v> e <b> no galego, no espanhol e no português do Norte de Portugal. Provavelmente, os grafemas <b> e <v> eram pronunciados ambos como bilabiais, tendo como distinção apenas o modo de articulação, sendo <b> uma oclusiva (como é hoje) e <v> uma fricativa [β]. Com o tempo, o grafema <v> passou de fricativa bilabial para fricativa labiodental [v], tal como é hoje, confundindo-se somente no espanhol, no galego, e no português do Norte de Portugal. No tempo em que os portugueses contactaram os antepassados dos atuais falantes de crioulos, contudo, a confusão permanecia. <b> e <v> possuíam uma leve diferença, apenas no modo de articulação, ou ainda, uma mesma pronúncia. Fósseis dessa variação podem ser encontradas ainda hoje, nas variações de ortoepia autorizadas por dicionários como

“assobio”/“assovio”, ou “basculhar”/“vasculhar” (Houaiss & Villar, 2001: 323, 410) ou ainda em variações discriminadas socialmente, como “taverna” e “vasculhante”, que se encontram desabonadas pelos dicionários, que dão preferência às formas “taberna” e “basculante”.

Essa variação persistiu até o final do século XV, sendo absorvida pelos crioulos africanos, isso explica a existência dos fonemas nas palavras são-tomenses: **bo** < vós, **'suba** < chuva, **bi'fi** < vestir, **ba'sola** < vassoura, **la'ba** < lavar. Em outros casos, o <v> permaneceu: **'vede** < verde, **navle'ga** < navegar (Ferraz, 1979: 35).

#### 4.2.1.5. Inventário fonológico das consoantes

**Tabela 13:** Sistema fonológico do português seiscentista.<sup>31</sup>

	bilabial	labiodental	alveolar	apicoalveolar (retroflexa)	alveopalatal	palatal	velar
Oclusiva	p b		t d		tʃ dʒ		k g
Nasal	m		n			ɲ	
Vibrante			r				
Tepe			ɾ				
Fricativa	[β]	f v	s z	ʂ ʐ		ʃ	
Lateral			l			ʎ	

Pares mínimos do português seiscentista:

- (9) /p/ **pato**  
/b/ **bato**
- (10) /β/ **cavo**  
/b/ **cabo**
- (11) /t/ **tente**  
/d/ **dente**

<sup>31</sup> É importante mencionar o fone [β] já que ele explica a variação entre [b] e [v] em são-tomense.

- (12) /tʃ/ **acha**  
/dʒ/ **haja**
- (13) /k/ **calo**  
/g/ **galo**
- (14) /m/ **gama**  
/n/ **gana**  
/ɲ/ **ganha**
- (15) /r/ **caro**  
/r/ **carro**
- (16) /f/ **faca**  
/v/ **vaca**
- (17) /s/ **cinco**  
/z/ **zinco**
- (18) /s/ **paço**  
/ʃ/ **passo**
- (19) /z/ **cozer**  
/z̃/ **coser**
- (20) /l/ **cala**  
/ʎ/ **calha**
- (21) /ʃ/ **taxa**  
/tʃ/ **tacha**

#### 4.2.1.6. O sistema vocálico

O sistema vocálico do português medieval, que serviu de superstrato aos crioulos do Golfo da Guiné, consistia de sete vogais orais (Silva, 1991: 51), a saber /i, u, ε, ɔ, e, o, a/. Oliveira (1536), em sua descrição da fonética do português do século XVI, menciona a existência de oito vogais orais. Mencionava a existência de um ‘a’ grande e um pequeno, um ‘e’ grande e outro pequeno, um ‘o’ grande e outro pequeno. Não confundia os fonemas com grafemas, afirmando haver em português apenas cinco figuras para oito vogais (Oliveira, 1536: cap. XII). As vogais grandes eram os fonemas abertos /a, ε, ɔ/, e as pequenas eram os fonemas fechados /ə, e, o/ (Coseriu, 2000: 35). A divergência entre a descrição de Oliveira (1536) e Silva (1991) está na menção, por parte de Oliveira (1536) da vogal medial fechada /ə/ que ele propunha que fosse grafada com um alfa grego. Essa vogal atualmente é um fonema em Portugal, fazendo a oposição vocálica existente entre a forma verbal “amamos” no presente do indicativo (grafada ‘amámos’ pela ortografia portuguesa, mas não na brasileira) e “amamos” no pretérito perfeito. Essa oposição fonológica não existia ainda no português do século XVI, segundo Silva (1991: 49). O fone [ə], hoje existente no português europeu, atualmente serve como alofone de [e] e de [a], desapareceu como fonema no português brasileiro, aparecendo apenas como alofone de [a] em palavras como “cama”, “cana”, etc. e, nos crioulos analisados, pôde ser encontrado apenas como alofone de /e/. Oliveira tinha consciência da existência do [ə], mas, dadas as limitações da época, não podia diferenciar fones e fonemas.

A oposição fonológica entre [ɐ] e [a], existente em línguas como o romeno [ˈkasɐ] “casa”, [ˈkasa] “a casa”, (Rauta, 1974: 61), não existe no português moderno e também não o existia no português seiscentista. O fone [ɐ] existe atualmente no português europeu e brasileiro, e aparece em posição final, mas é inexistente nos quatro crioulos do Golfo da Guiné. Oliveira (1536) não mencionava a existência dessa vogal, assim como Silva (1991) também não o menciona. Isso serve como mais uma evidência de que a formação do sãotomense se deu em épocas muito remotas e de que as influências do português moderno não atingiram esse aspecto fonético.

O sistema vocálico oral do português do século XVI, resumidamente, apresentava sete fonemas vogais [i, u, ε, ɔ, e, o, a] e um fone vogal [ə].

A questão da existência das vogais nasais em português, tanto no seiscentista quanto no atual, é muito controversa. Camara Jr. (1998: 46-47, 58-60) já mencionava que as vogais nasais não existem em português, de um ponto de vista estrutural, e que só podem ser vistas como vogais nasalizadas por consoantes nasais contíguas na posição de coda. Oliveira (1536), Buescu (1975), Silva (1991), Coseriu (2000) e Teyssier (2004), ao contrário, assumem que havia cinco vogais nasais no português seiscentista. Dessa forma, podemos resumir o sistema vocálico do português seiscentista da forma como se segue.

**Tabela 14:** sistema vocálico I.

<b>Vogais orais</b>			
	Anteriores	Central	Posteriores
Altas	i		u
Médias-altas	e		o
Médias-baixas	ɛ		ɔ
Baixas		a	

**Tabela 15:** sistema vocálico II.

<b>Vogais nasais</b>			
	Anteriores	Central	Posteriores
Altas	ĩ		ũ
Médias-altas	ẽ		õ
Médias-baixas	-		-
Baixas		ã	

#### 4.2.1.7. O ditongo [ɣũ]

O atual ditongo [ãw] do português do Brasil e [ɣũ] do português europeu, ambos grafados atualmente como <ão> em substantivos e <am> nos verbos no pretérito perfeito e imperfeito, eram pronunciados, no português arcaico, como [õ] (Coutinho, 1976: 66 e Ferraz, 1979: 31).

Do surgimento da língua portuguesa até meados do século XV, havia uma variação fonética entre [õ] e [ɔ̃]<sup>32</sup>. Entre a metade do século XV e a metade do XVI, a fusão dessas duas variantes se completou (Ferraz, 1979: 31), restando no português europeu e brasileiro atuais apenas a forma ditongada. Já em galego moderno, espanhol e nos crioulos de base portuguesa, tal ditongo não existe.

Lucchesi & Baxter (2006) e Oliveira (2007) apresentam frases do português de Helvécia, na Bahia, falado no século XXI: “A véia Veronca ã é mãe dela non. / Quando chega lá, eles ã queria que eu sai ã.” Em que podemos observar uma monotongação nasal do ditongo português [ãw] > [õ]. Se é possível afirmar que ocorreu uma monotongação por influência fonéticas africanas, também é possível afirmar que houve, na verdade, uma conservação da pronúncia portuguesa do século XVI, época da colonização portuguesa no Brasil. Na carta de Caminha, já citada no início da seção 3, podemos observar as formas ortográficas <nõ> para o advérbio de negação <não>, e a forma ortográfica <capitam> para o atual <capitão>, embora o ditongo já pudesse ser pronunciado, enquanto a ortografia ainda estava contaminada pelas formas antigas. Porém, na primeira edição de *Os Lusíadas*, em 1572, já aparece a forma ortográfica <não> (Camões, 1572, c. I, e. 5, v. 2). Para Ferraz (1979: 31), a permanência em sãotomense da variante medieval [õ] é uma prova de que a formação do sãotomense se deu em épocas muito remotas. Partindo desse pressuposto e, levando-se em consideração que para Teyssier (2004: 56) o ditongo já era geral em 1500, a formação do crioulo deve-se ter dado entre o final do século XV e meados do século XVI.

Nos crioulos de base portuguesa não existe o ditongo nasal [ãw], conforme já foi dito. As palavras portuguesas terminadas com esse ditongo terminam em sãotomense ora com [ã] ora com [õ], como em **lungwa san'tome** “língua de São Tomé”, **kasõ** “caixão”, **mõ** “mão” (Ferraz, 1979: 8, 20, 22); em principense com /ã/ ou /ɛ/ como em **fɔ'gã** “fogão” e **fal'kɛ** “falcão”.

#### 4.2.1.8. O ditongo [ow]

O ditongo [ow], em palavras como **couro**, **roupa** e **amou**, num processo fonético vindo do Sul de Portugal (Teyssier, 2004: 63), passou de [ow] a [o], um monotongo

---

<sup>32</sup> Ferraz (1979) representava tal fonema nasal [ɔ̃] com o til subscrito e não sobrescrito. Decidimos manter a notação do autor.

comum em todo o Brasil, onde se pronuncia c[o]ro, r[o]pa, am[o], em todas as variantes diastráticas. Esse processo começou em Portugal no século XVII, marginalizando o uso ditongado, ainda vivo no norte de Portugal. Em algumas palavras, contudo, o [ow] se transformou em [oj]: touro > toiro; ouro > oiro; cousa > coisa (Teyssier, 2004: 63).

Embora a monotongação só tenha se dado no século XVII, no sãotomense o ditongo [ow] se monotongou em [o] já na formação do crioulo (séc. XV), como: **dotolo** < doutor, **otlo** < outro (Ferraz, 1979: 10, 40). O ditongo [oj], vindo ou não de um [ow], transformou-se também em [o]: **dodo** < doido, **dosu** < dois, **oto** < oito (Ferraz: 1979: 33, 72, 73).

#### 4.2.1.9. O ditongo [ej]

A monotongação do ditongo [ej], para Teyssier (2004: 77) ao contrário da monotongação do [ow], ocorreu em data difícil de se determinar, e não foi incorporada à norma do português contemporâneo, por não ter atingido a cidade de Lisboa, local da variante de prestígio. Pelos textos teatrais do século XVIII, mencionados por Teyssier (2004: 77), calcula-se que, pelo menos naquele século, a monotongação já era comum.

No século XIX, inovações fonéticas nas faixas portuguesas onde não havia o monotongo levaram o ditongo [ej] a se transformar em [aj] e o ditongo [ẽj] a se transformar em [ãj]. Assim, **peito** > p[aj]to, tamb[ẽj] > tamb[ãj]. No poema “O menino da sua mãe”, de Fernando Pessoa, pode-se observar uma rima perfeita entre ‘tem’ e ‘mãe’. Tal alteração fonética, contudo, nunca foi copiada por brasileiros em sua fala quer formal ou informal, eis um dos traços distintivos da fala europeia.

No Brasil, apesar de não se transformar em [ãj], o ditongo [ej] se monotongou antes das consoantes fricativas palatais [ʒ] e [ʃ], da oclusiva velar [g], e da vibrante [r]. “Beijo”, “peixe”, “manteiga”, “cheiro”: [ˈbezu], [ˈpefi], [mãˈteɣɐ], [ˈʃeru]. Nas demais situações, continuou pronunciado como ditongo: “teima”, “reino”, “beijo”, “leite”<sup>33</sup>, “seiva”, “Deise”, e, como o ditongo é inexistente antes das oclusivas bilabiais [b] e [p] e da velar [k] no vernáculo, aparece nos empréstimos: “playboy”, “milk shake”, “video tape”, sem sofrer redução na fala popular. É importante notar, ainda, as ditongações que ocorrem com as palavras como “bandeja”, “caranguejo” e “tempero”. Mais

---

<sup>33</sup> O mesmo ocorre no português brasileiro do Nordeste: peito [pejtʃu] em que não há monotongação, pois a presença do [ʃ] é após um [t].

acuradamente, nota-se que tal fenômeno ocorre por hipercorreção. Ciente do desprestígio social da monotongação, diversos falantes brasileiros tendem a ditongar os monotongos antes das palatais, como em “ameixa” [a'meʃɐ] e “peixe” [ˈpɛʃi], o que os leva, por analogia, a ditongar vogais simples, tal como comumente ocorre na literatura de cordel, com formas “não se aveixe” no lugar de “não se avexe”. Gramáticas normativas como Nicola & Infante (1994: 59) registram como “erros” de ortoepia, as ditongações de “bandeja” e “caranguejo”, e as monotongações de “cabeleireiro”. Confronte-se ainda a pronúncia carioca corrente de [ˈdowzi] para ‘doze’, e as confusões de pronúncia entre os verbos ‘pousar’ e ‘posar’.

No crioulo sãotomense, o ditongo português [ej] se monotonga em qualquer situação. Quando a palavra portuguesa termina em [u], o [ej] se transforma em [e], quando a palavra portuguesa termina em [a], o [ej] se transforma em [ɛ]. Primeiro > **plu'me**, peito > **petu**, pedreiro > **pedlelu**, e ainda: seis > **sefi** (Ferraz, 1979: 73); cadeira > **ka'dela**, poeira > **pwela**, deixar > **deʃa** (Ferraz, 1979: 31). Ferraz parece supor que a pronúncia [aj] para o ditongo [ej] já era comum para os portugueses desde o século XVI<sup>34</sup>, o que contradiz Teyssier (2004: 78).

---

<sup>34</sup> Teyssier (2004: 78) calcula que a transformação de [êj] em [ãj] tenha ocorrido no século XIX. Já a de [ej] para [aj], afirma ser um fato consumado na segunda metade do século XVIII, embora creia que suas primeiras manifestações fossem mais antigas. Cabe-nos questionar quão mais antigas sejam. Ferraz (1979) pressupunha que a transformação já estivesse completa no século XV, o que parece exagerado.



## 5. FONÉTICA E FONOLOGIA DO SÃOTOMENSE

Para a descrição fonológica do sãotomense, utilizou-se como base teórica a obra de Ferraz (1979). Como corpus principal, usou-se a lista-Swadesh de Graham & Graham (2004), cujos dados foram cruzados com Valkoff (1966) e Mane (2007).

Dos 29 informantes que cederam informações para que Graham & Graham (2004) montassem o seu corpus, apenas nos interessa explicitar, primeiramente, dois que forneceram dados para as línguas estudadas neste trabalho.

Os informantes do sãotomense são:

ST1: uma professora primária, nascida em Trindade, São Tomé, com 51 anos de idade e 9 anos de estudo, além dos três anos de treinamento para o magistério (equivalendo no Brasil ao Ensino Fundamental e o Ensino Médio Normal, para formação de professores, logo, uma informante ex-normalista com ensino médio completo).

ST2: um secretário administrativo nascido em Água Grande, São Tomé, com 30 anos de idade e 5 anos de estudo (equivalendo no Brasil aos dois primeiros ciclos do Ensino Fundamental, logo, um informante que estudou até a “4ª série”).

Há de se ressaltar ainda que, pelas profissões exercidas pelos informantes, muito provavelmente, são falantes fluentes de português e utilizam essa língua com frequência, embora em Graham & Graham (2004) não haja menção sobre isso.

### 5.1. Inventário fonético das consoantes

O sistema fonético do sãotomense não foi descrito por Ferraz nem por Mane, que se mantiveram na descrição fonológica. Achamos importante essa primeira descrição para notar, mais profundamente, a influência de línguas africanas no crioulo estudado neste trabalho.

O inventário de fones do sãotomense é, como ocorre em todas as línguas, mais rico que o de fonemas. Contudo, a verdadeira riqueza de detalhes do inventário fonético se mostra ao constatarmos que, há entre os fones do sãotomense, sons existentes apenas em línguas africanas, como os cliques, ou mais comuns a elas que a quaisquer outras línguas, como as consoantes não-pulmônicas (Heine, 2006). Inexistentes no português, oito dos 25 fones consonantais listados nas tabelas 16, e os três fones listados na tabela 17, só podem ser, por eliminação, de origem africana.

**Tabela 16:** Consoantes pulmônicas.

	bilabial	lábio-dental	alveolar	pós-alveolar	retroflexa	palatal	velar	uvular	glotal
Oclusiva	p b		t d				k g	q	ʔ
Nasal	m		n			ɲ	ŋ	ɴ	
Trill			r						
Tepe (ou flepe)			ɾ						
Fricativa		f v	s z						
Lateral fricativa				ɬ ɮ					
Flepe lateral			ɭ						
Aproximante						j			
Lateral aproximante			l		ɭ				

**Tabela 17:** consoantes não-pulmônicas.

ɓ	Implosiva bilabial sonora
ɗ	Implosiva dental/alveolar
	Clique alveolar lateral

Em seguida, daremos os exemplos de fones. Após cada um deles, damos o exemplo em são-tomense e a tradução. Os dados podem ser conferidos na lista-swadesh de Graham & Graham (2004).

[ʔ] – [ʔuwe] – “olho” – oclusiva glotal. Não existe em português. “Tal fonema se realiza com uma pequena interrupção da corrente de ar, seguida por um súbito relaxamento da glote” (Navarro, 1999). Funciona, geralmente, como fone inicial nas palavras que, no português, iniciam por vogal.

[ɓ] – [ʔboka] – “boca” – bilabial implosiva vozeada. Muito comum em línguas africanas (Heine, 2006) mas não existe nos principais substratos do são-tomense, o quicongo e o bini (Ferraz, 1979: 49). Funciona como alofone de [b]. O mesmo vocábulo foi pronunciado [ʔboka] por outro falante.

[ŋ] – [lʉŋg<sup>w</sup>a] – “língua” – nasal velar vozeada. Em sãotomense só aparece antes de velares. Devido a esse fato, provavelmente, Mane (2007) cita fonemas não mencionados por Ferraz (1979), o grupo das pré-nasais: bilabiais, alveolares e velares, estas últimas sendo /ŋk/ e /ŋg/ (Mane, 2007: 117 e 128). Os grupos [nk] e [ng] existem em quicongo (cf. tabela 5).

[l̥] – [kabel̥] – “cabelo” – lateral aproximante retroflexa. Ocorre em sãotomense como um alofone do [l] das palavras em português, ou de palavras originalmente com [r] – marido [mal̥ũ] ~ [mar<sup>h</sup>idu].

[q] – [qʰoŋ qʰo] – “pescoço” – oclusiva uvular surda. A palavra “pescoço”, registrada por Ferraz (1979: 98) como klõ-<sup>h</sup>klõ, é registrada de duas formas por Graham & Graham (2004), devido à existência de dois falantes de origens geográficas diferentes [qʰoŋ qʰo] e [kloŋklo]. Na primeira das variantes, são utilizadas duas uvulares e uma alveolar, na outra, os fonemas comuns à língua portuguesa, com exceção da nasal velar.

[ʎ] – [qʰoŋ qʰo] – “pescoço” – lateral alveolar surda. Vide exemplo anterior.

[N] – [qʰoŋ qʰo] – “pescoço” – nasal uvular. Usada como homorgânica no exemplo [qʰoŋ qʰo] por ser [q] também uma consoante uvular.

[r] – [ma<sup>h</sup>ridu] – “marido” – flepe alveolar. Extremamente raro em sãotomense, aparece como alofone de [l].

[l̥] – [kabl̥a] – “cabra” – flepe lateral alveolar. A forma [kabl̥a] aparece nos dois informantes. Aparece em algumas palavras do sãotomense como um alofone das palavras cognatas em outros crioulos do Golfo da Guiné, onde usam o tepe alveolar [r].

[d̥] – [l̥<sup>h</sup>idu] – “piolho” – implosiva dental/alveolar. A forma [l̥<sup>h</sup>idu] aparece para os dois informantes. Ocorre em algumas palavras como alofone de [dʒ] em relação a outros crioulos do Golfo da Guiné.

[j] – [ʔija] – “lenha” – aproximante palatal. Na verdade este é um fonema, não um fone, mas como não existe par mínimo contrastivo, trouxemos este fonema para o inventário. Ferraz (1979: 22) transcreve galinha como [ngaʎna] e Graham & Graham (2004) transcrevem como [ᵑgaʎjá] com ‘i’ nasal seguido de ‘j’. Esta é a estrutura que existia em português medieval em palavras como <v̄io>, havendo, foneticamente, provavelmente um [v̄ijo], tendo essa aproximante palatal posteriormente se nasalizado, por conta da vogal nasal que a precede, transformando-se numa nasal palatal [ɲ]. Já que a palavra galinha vem da evolução: galina > galĩa > galinha, teria vindo a forma sãotomense do português arcaico?

[k̠] – [ʎi] – “rir” – lateral fricativa alveolar vozeada. Nos exemplos encontrados, equivale ao “r” inicial em português. A forma [ʎi] ‘rir’ aparece para os dois informantes. Mas também “ladrar” [ʎad̠a].

[||] – [n||da] – “andar” – clique alveolar lateral. Os cliques são fones/fonemas que, ao que tudo indica, só existem em línguas africanas, mais precisamente nas do tronco khoisan (Heine, 2006) como as línguas dos bosquímanos da Namíbia. Em sãotomense, aparece como um fone que não substitui qualquer vogal portuguesa. Exemplo: andar [n||da], mesma palavra em anobonês e angolar, mas em principense são [ʔanda] ~ [ʔinda]. Mane (2007: 126) não menciona o clique, e cita [nda] como andar, numa lista em que tenta demonstrar a existência de fonemas pré-nasais. Os cliques são raros em sãotomense, dos 170 vocábulos listados, apareceu em apenas dois, [n||da], andar, e [n||tela] enterrar. Apareceram, pelo que se pôde observar, devido a uma suarabácti. O fone funcionou como uma vogal epentética para desfazer um encontro consonantal em onset. Como os dois exemplos foram os únicos com cluster [nt] em onset, não é possível dizer que os cliques sejam consistentes foneticamente. Ideal seria um corpus maior com mais exemplos de palavras com cluster [nt] em onset para tirarmos conclusões. Note-se ainda que em [n||tela], última palavra da lista de Graham & Graham (2004), a sílaba tônica é a segunda, dando a entender a formação de uma sílaba n + clique, daí deduzirmos serem os cliques relevantes em sãotomense.

## 5.2. Inventário fonológico das consoantes

**Tabela 18:** Inventário fonológico (baseado em Ferraz, 1979).

	bilabial	labiodental	alveolar	Postalv.	palatal	velar
Oclusiva	p b		t d		tʃ dʒ	k g
Nasal	m		n		ɲ	
Vibrante			r			
Fricativa		f v	s z	ʃ ʒ		
Aproximante					j	
Lateral	w		l		ʎ	

Exemplos de pares mínimos (Ferraz, 1979: 21), traduzido e com alfabeto fonético atualizado:

- (22) /p/    **pa'li**        parir  
           /b/    **ba'li**        pântano
- (23) /t/    **'tadzi**        tarde  
           /d/    **'dadzi**        idade
- (24) /k/    **'kãsu**        asma  
           /g/    **'gãsu**        gancho
- (25) /f/    **flɛ'ga**        esfregar  
           /v/    **vlɛ'ga**        curvar (envergar?)
- (26) /s/    **ka'sa**        caçar  
           /z/    **ka'za**        casar
- (27) /ʃ/    **'maʃi**        mais  
           /ʒ/    **'maʒi**        mas

- |      |      |               |                    |
|------|------|---------------|--------------------|
| (28) | /tʃ/ | <b>mã'tʃa</b> | saudação           |
|      | /dʒ/ | <b>mã'dʒa</b> | madrinha           |
| (29) | /m/  | <b>mõ</b>     | mão                |
|      | /n/  | <b>nõ</b>     | nós                |
| (30) | /m/  | <b>nga'ma</b> | gamela             |
|      | /ɲ/  | <b>nga'ɲa</b> | galinha            |
| (31) | /n/  | <b>ba'na</b>  | sacudir (abandar?) |
|      | /ɲ/  | <b>ba'ɲa</b>  | bainha             |
| (32) | /l/  | <b>mo'le</b>  | morrer             |
|      | /d/  | <b>mo'de</b>  | morder             |

Obs.: São raros os fonemas /ɲ/, /ʎ/ (Ferraz, 1979: 22) e /r/ (Ferraz, 1979: 37).

### 5.3. Inventário fonético das vogais

Foneticamente, o sãotomense, pelos dados levantados por Graham & Graham (2004), possui nove vogais orais. Além das sete fonológicas, dois fones /ə/ e /ɪ/. O primeiro funciona como alofone de /a/ em posição final: pena /<sup>h</sup>pɛnə/ ~ /<sup>h</sup>pɛna/ e faca /<sup>h</sup>fakə/ ~ /<sup>h</sup>faka/. O segundo, como alofone de /i/, como se nota nos exemplos: beber /<sup>h</sup>bɪbɛ/ ~ /<sup>h</sup>bibi/. Embora em português existam diferenças fonológicas entre vogais abertas e fechadas, e Ferraz (1979), Rougé (2004) e Mane (2007) advoguem pela existência de sete vogais orais em sãotomense, do ponto de vista fonético parece haver variação diatópica entre os falantes, possibilitando a entrada de algumas vogais no campo dos arquifonemas. Entre os dois informantes de sãotomense, há constante variação entre abertas e fechadas: o pó /<sup>h</sup>ʔopo/ ~ /<sup>h</sup>ʔɔpɔ/, nove /<sup>h</sup>novɛ/ ~ /<sup>h</sup>nɔvɛ/, fome /<sup>h</sup>fomi/ ~ /<sup>h</sup>fɔmi/ ou ainda, entre outras vogais: nuvem /<sup>h</sup>novi/ ~ /<sup>h</sup>nuvɛ/ e ainda casos de

variação em que, induzidos a dizer as palavras “longo” e “alto”, um falante pronuncia /lɔŋgo/ para os dois casos, enquanto outro diz /lɔŋgo/ para uma e /luŋgo/ para outra.

#### 5.4. Inventário fonológico das vogais

O sistema fonológico do sãotomense foi descrito por Luiz I. Ferraz, em sua tese de doutoramento pela Universidade de Johannesburgo, África do Sul. O autor emprega um alfabeto fonético biunívoco, usando um símbolo gráfico para cada fonema, e representando um mesmo fonema por um único símbolo gráfico. O autor marca todas as vogais nasais graficamente com til, já que não parece considerar a existência de consoantes nasalizantes. Como exemplo, uma palavra do sãotomense como “quente” é escrita [kêci], quando poderia ser feito atualmente: [kɛntʃi], [kɛnci], ou ainda, da forma como preferimos neste trabalho: [kɛ̃tʃi].

Ferraz descreve a fonologia do sãotomense nas páginas 19 a 57, e chega a propor uma ortografia (páginas 55 a 57) para o sãotomense, já que, embora a língua tenha mais de quatrocentos anos, não há uma ortografia unificada e nem mesmo uma oficial. Apesar de propor essa ortografia, ele não a utiliza na maior parte do livro.

Há doze vogais fonológicas no sãotomense, tal como descreve Ferraz (1979: 20), e como pode ser visto nos quadros a seguir.

**Tabela 19:** inventário das vogais orais do sãotomense.

Vogais orais			
	Anteriores	Central	Posteriores
Altas	i		u
Médias-altas	e		o
Médias-baixas	ɛ		ɔ
Baixas		a	

**Tabela 20:** inventário das vogais nasais do são-tomense.

Vogais nasais			
	Anteriores	Central	Posteriores
Altas	ĩ		ũ
Médias-altas	ẽ		õ
Médias-baixas	-		-
Baixas		ã	

Exemplos de pares mínimos (Ferraz, 1979: 20), traduzido e com alfabeto fonético atualizado:

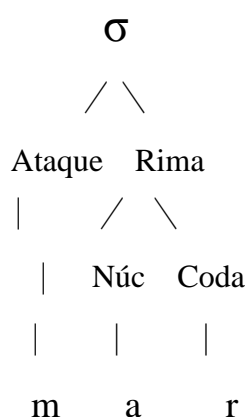
- (33) /a/    **'basa** onda  
           /ã/    **'bãsa** costela
- (34) /e/    **te**    ter  
           /ẽ/    **tẽ**    também
- (35) /i/    **'fita**    fita  
           /ĩ/    **'fita**    coleção
- (36) /o/    **ka'so**    cachorro  
           /õ/    **ka'sõ**    caixão
- (37) /u/    **'fudu**    limpo  
           /ũ/    **'fũdu**    fundo
- (38) /i/    **'pitu**    apito  
           /e/    **'petu**    peito  
           /ɛ/    **'petu**    perto
- (39) /u/    **ku'je**    colher (verbo)  
           /o/    **ko'je**    escolher



- (40) /o/    **'sotʃi** soprar, ventar  
       /ɔ/    **'sotʃi** sorte

### 5.5. A estrutura silábica

Podemos considerar as sílabas como unidades sonoras compostas de três elementos básicos: ataque, núcleo e coda. O ataque é o elemento inicial da sílaba, como o “p” da palavra “pato”, ou o “pr” da palavra “prato”. O núcleo é o ponto máximo central de sonoridade, o elemento principal, geralmente uma vogal, sobre a qual as consoantes do ataque e da coda se apoiam. A coda é o elemento final da sílaba, geralmente uma consoante ou consoantes que fecham a estrutura silábica, como o “z” da palavra portuguesa paz. Clements & Keyser (1983) criaram um sistema estrutural para o estudo das sílabas. Segue um exemplo do sistema estrutural com um monossílabo português “mar”.



As sílabas podem ser compostas por apenas uma vogal, como a primeira sílaba da palavra portuguesa “ato”, em que “a” é formada de um ponto máximo (vogal) isolada, sem qualquer consoante em apoio. Já a segunda sílaba da palavra, “to”, é composta de uma vogal “o”, pronunciada [u], com uma consoante que se apoia nesse ponto máximo sonoro.

Dos elementos básicos da sílaba, o ataque (também chamado de onset) e a rima, apenas a última é obrigatória. Na rima, o núcleo da sílaba é de presença obrigatória, já a coda é opcional. No caso de sílabas terminadas em vogais, como as da palavra “casa”, dizemos que são sílabas simples, leves ou abertas, por não possuírem coda. Já as sílabas terminadas em consoante são chamadas de complexas, pesadas ou fechadas. Em

português só podem permanecer em posição de coda as consoantes S, Z, R, L, N, em suas diversas realizações dialetais, como as fricativas surdas [s] ou [ʃ]: casca [ˈkaskɐ] em São Paulo, ou [ˈkafkɐ] no Rio de Janeiro; como as fricativas sonoras [z] ou [ʒ]: rasgo [ˈRazgu]<sup>35</sup> em São Paulo, ou [ˈRazgu] no Rio de Janeiro; a fricativa velar sonora [ɣ]: carga [ˈkayɣɐ] no Rio de Janeiro ou a fricativa glotal vozeada [ɦ]: carga [ˈkaɦɣɐ] em Belo Horizonte, etc. (Exemplos de Silva, 1999: 37, 38). No português do Brasil, ao contrário do português europeu, não existem consoantes oclusivas em posição de coda. No Brasil, uma nova sílaba é criada por epêntese, ou seja, uma vogal é acrescentada no local da coda da sílaba anterior, e uma nova sílaba é criada: “dogma” [ˈdɔgmɐ] > [ˈdɔgimɐ], “pneu” [ˈpnew] > [peˈnew]. Essas pronúncias são criadas devido às regras fonotáticas do português brasileiro, que não permitem o travamento de sílabas por oclusivas. Devido aos mesmos motivos, a hipercorreção leva muitos falantes de português brasileiro a tentarem pronunciar [adʒviˈɲa] em vez de [adʒiviˈɲa], forçando assim consoantes fechadas em palavras com consoantes abertas. Gramáticas normativas costumam registrar essas hipercorreções como “erros” de ortoepia (Nicola & Infante, 1994: 59).

As sílabas compostas por apenas uma vogal são sílabas V. As compostas por vogal e consoante são CV, e assim por diante. As sílabas CV (consoante + vogal) são as mais comuns e existem em todas as línguas do mundo, são chamadas sílabas universais. Nos casos dos ditongos, a semivogal (ou glide) também pode ser representada por V, daí termos sílabas VV, como na palavra “oi”. Encontros consonantais, chamados também de clusters, são representados por CC, daí termos sílabas CCV, como em “pra.to”. Quando um segmento se apresenta no início da primeira sílaba de uma palavra, enfim, no início da palavra, chamamos ataque absoluto; quando um segmento se apresenta no fim de uma sílaba que também é a última da palavra, ficando esse segmento no final da palavra, chamamos coda absoluta.

Em seu estudo do padrão silábico do sãotomense, Mane (2007: 135) utilizou, para o sãotomense, uma lista de 510 palavras que, juntas, somavam o *corpus* de 1.167 sílabas. Para as informações de frequência, o cálculo realizado foi número de sílabas de cada tipo encontrada multiplicada por 100 e dividido por 510 (número de palavras) chegando a um total de frequência de 228,8235%. Para o nosso trabalho, vamos nos

---

<sup>35</sup> [R] é um arqui fonema que representa os fones [h] de Belo Horizonte, [χ] no Rio de Janeiro e [r] em Portugal e no português caipira.

basear num sistema de frequência centesimal, e para atingir a soma de 100%, dividiremos a quantidade de sílabas encontradas por 11,67, já que o corpus é de 1.167 sílabas, e  $1.167 / 11,67 = 100$ . Logo, neste trabalho, o valor em porcentagem encontrado na coluna frequência pode ser analisado, estatisticamente, como a quantidade de sílabas de tal tipo encontradas no sãotomense.

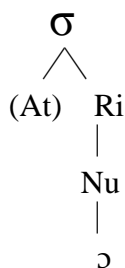
**Tabela 21:** frequência das sílabas.

Tipo	Sãotomense	
	exemplo	frequência
CV	[ka.ba] acabar	80,90%
V	[ɔ.la] hora	7,37%
CCV	[li.vlu] livro	6,60%
CVC	[poʃ.ta] apostar	3,69%
CCCV	[ʃtla.da] estrada	1,11%
VC	[iʃ.ka] isca	0,25%
CCVC	[flɛs.ku] fresco	0,08%

As sílabas do padrão CV, existentes em todas a línguas do mundo, são as mais comuns em sãotomense: cerca de 80% de todas as sílabas da língua. Cerca de 7% são de sílabas compostas por apenas uma vogal. As terceiras mais comuns, CCV, aparecem em 6,6% dos casos. Juntas, essas três estruturas silábicas correspondem a quase 95% de todas as sílabas em sãotomense.

A estrutura dos três tipos mais comuns de sílabas em sãotomense podem ser apresentadas como se segue:

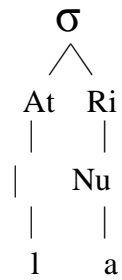
- (41) 1. V  
       ɔ.la - hora



As sílabas do tipo V, compostas apenas por uma vogal, que são 7%, são mais comuns em principense, cerca de 12%.

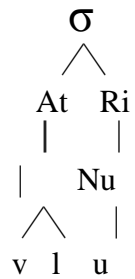
(42) 2. CV

o.la - hora



(43) 3. CCV

li.vlu - livro



A seguir daremos o quadro das regras fonotáticas do são-tomense, organizado com base nas informações de Ferraz (1979: 26-28) e Mane (2007).

**Tabela 22:** regras de formação das sílabas.

<b>Padrão</b>	<b>Regras fonotáticas</b>	<b>Exemplo</b>	<b>Tradução</b>
CV	Qualquer consoante ou vogal é permitida <sup>36</sup> .	'tɛ.la	terra
V	Ocorre quase sempre em início de palavra: 1. A palavra portuguesa tinha a primeira sílaba tônica; 2. A palavra portuguesa era monossílaba, o artigo se fundiu à primeira vogal.	'awa ɔ'mali	água o mar
C <sub>1</sub> C <sub>2</sub> V	C <sub>1</sub> é qualquer consoante C <sub>2</sub> é [w] ou [l]  ou C <sub>1</sub> é uma homorgânica nasal	tla'ba, bwe  nda	trabalhar, boi  andar
C <sub>1</sub> VC <sub>2</sub>	V é sempre oral e C <sub>2</sub> é fricativa palatal  ou C <sub>2</sub> é sempre uma homorgânica nasal	'mɔʃ.ka  'sum.bu	mosca  chumbo
C <sub>1</sub> C <sub>2</sub> C <sub>3</sub> V	C <sup>1</sup> é sempre uma fricativa surda C <sub>2</sub> é sempre /t/ ou /k/ C <sub>3</sub> é sempre /l/	'ʃtlɛ.ka ʃkle've	cercar escrever
C <sub>1</sub> C <sub>2</sub> VC <sub>3</sub>	Um único caso encontrado. Ver informações abaixo.	'fleʃku	fresco
VC	C é sempre uma fricativa palatal ou uma nasal	'iʃka	isca

A sílaba CCCV só ocorre em início de palavra. Os raros casos em que ocorre no meio de palavra, ocorre após vogal nasal, como em mɛ.ʃtla'sõ (menstruação).

Ferraz (1979) não menciona a existência das sílabas CCVC e VC, que Mane (2007) lista em seu trabalho. Mane só encontrou uma única sílaba CCVC num universo de 510 palavras. Pela palavra encontrada: fleʃ.ku (fresco) nota-se que as regras fonotáticas das sílabas CCV e CVC foram cruzadas: C<sub>2</sub> é um [l], regra do CCV, V é uma vogal oral e a consoante em coda é uma fricativa palatal, regra do CVC. A palavra fleʃku é mencionada por Ferraz (1979: 34). Pode ser que ele a considerasse como fle.ʃku e não como fleʃ.ku, assim como mla.ʃka. de qualquer modo, o padrão silábico foi incluído na tabela 22. No caso das sílabas VC, Mane só encontrou três casos, em 510 palavras. Pelo exemplo citado, notamos que as regras mais uma vez são derivadas das primeiras. A consoante em coda é uma fricativa palatal, aparentemente o único tipo de consoante permitida nessa posição.

<sup>36</sup> A lateral palatal [ʎ] não aparece em início de palavra, a nasal velar [ŋ] não aparece em início de sílaba em meio de palavra.

## 6. METAPLASMOS

Metaplasmos são modificações fonéticas que as palavras sofrem em sua evolução. Podem ocorrer sincronicamente (ao mesmo tempo) ou diacronicamente (ao longo da evolução da língua). Originalmente eram usados para o estudo sincrônico da língua, classificando e nomeando traços de alterações fonéticas entre as diversas variantes, como, por exemplo, inimigo~imigo, perla~pérola, posteriormente, passou a ser usado em linguística histórica para nomear as alterações fonéticas ocorridas ao longo da evolução linguística, como na troca dos fonemas finais do latim <semper> para o português <sempre> (Camara Jr. 1968: 263).

Para os exemplos de metaplasmos ocorridos do latim para o português, usaremos os dados de Coutinho (1976: 142-149). Para os exemplos ocorridos do português para o são-tomense, usaremos Ferraz (1979). Duas distinções precisam ser feitas sobre os metaplasmos ocorridos do latim para o português e do português para o são-tomense.

Em primeiro lugar, não houve crioulização do latim, para que houvesse a passagem do léxico latino para o português. As palavras latinas sofreram, ao longo de séculos de transformações fonéticas, alterações para a criação do léxico português moderno. Em são-tomense, o processo foi mais rápido. Em menos de um século, o léxico português foi emprestado para o provável pidgin que serviu de base para o crioulo são-tomense, e as alterações fonéticas se deram quase que instantaneamente. De acordo com os dados de Ferraz (1979), elementos como a presença do fonema [õ] no lugar de [ãw] levam à conclusão de que o surgimento do crioulo são-tomense se deu antes da metade do século XVI, ou seja, como a colonização de São Tomé se deu a partir de 1485 (cf. Síntese Histórica), a crioulização e todas as alterações fonéticas se deram em menos de um século.

Em segundo lugar, há de se enfatizar que os metaplasmos ocorridos do latim para o português, ao longo de muitos séculos, se deram por causa de fatores externos ou por causa da deriva linguística, regras naturais da evolução das línguas levou à transformação fonética dos vocábulos latinos até o português arcaico, e do português arcaico até o português moderno. Já em são-tomense, provavelmente as línguas africanas de substrato afetaram positivamente essas alterações. Obviamente, da criação do são-tomense entre os séculos XV e XVI até a publicação da gramática de Ferraz (cinco séculos depois), algumas alterações fonéticas devem ter ocorrido por conta da deriva das línguas ou da influência posterior da língua portuguesa. Contudo, alguns traços

idiossincráticos dos metaplasmos são-tomenses, como o lambdacismo, a transformação do [r] em [l] com em ferro > <sup>l</sup>felu, ocorreram, muito provavelmente, já no processo de criouliização, devido às influências fonológicas de substrato, isto é, devido à inexistência dos fonemas [r] e [ɾ] em quicongo. Os africanos falantes dessa língua, os bakongo, ao aprenderem as palavras portuguesas, imprimiam a elas o seu “sotaque” característico.

O estudo dos metaplasmos em são-tomense será importante, não apenas para uma maior compreensão da fonologia do crioulo estudado, mas também para compreender o percurso de alterações fonéticas ocorridos no processo de empréstimos do português, como também as influências do substrato africano.

### 6.1. A estrutura das sílabas e as regras fonotáticas

Pode-se atribuir a maior parte dos metaplasmos do são-tomense à organização interna de sua estrutura silábica. Antes de apresentarmos os metaplasmos, seria imprescindível levar em consideração a estrutura silábica e as regras fonotáticas do são-tomense que foram apresentados na seção 5.5.

As regras fonotáticas são restrições sequenciais dos fonemas nas sílabas. Em português, sequências de consoantes como [sp] e [rt] não ocorrem em posição nenhuma em português (dentro de uma mesma sílaba). Sílabas não podem ter uma consoante fricativa alveolar em onset, seguida de outra consoante como em ‘sport’. Devido a essas regras de estruturação das sílabas – regras fonotáticas – ocorrem a maior parte dos metaplasmos. No caso do português, ao tomar a palavra inglesa por empréstimo, ocorreu o acréscimo de um fonema no início da palavra (prótese) para desfazer o cluster (encontro consonantal) não permitido em português em posição de onset (e qualquer outra posição na mesma sílaba). Como outra regra fonotática não permite o travamento de sílaba por qualquer oclusiva, como o [t], houve o acréscimo de outro fonema vocálico ao fim da palavra (paragoge), transformando o monossílabo inglês no trissílabo português “es.por.te”.

Em quicongo, todas as sílabas são leves, isto é, fechadas em vogal e nunca em consoante (Coene, 1960: 1; Lumwamu, 1973: 25). O elemento inicial é sempre uma consoante simples ou pré-nasal [mb], [ng], etc. assim como na maior parte das línguas bantas, as consoantes são sempre CV, as palavras são em sua maioria dissílabas CV.CV (Ferraz, 1979: 47). Em bini, todas as sílabas também são leves (Ferraz, 1979: 48). As

sílabas V geralmente só ocorrem em início de palavra, e em sua maioria substantivos. O sãotomense parece seguir essa estrutura africana, tal como se pode notar pelos metaplasmos que ocorreram durante o processo de empréstimos.

Os metaplasmos podem ser: por permuta ou transformação, por aumento, por subtração e por transposição.

## 6.2. Metaplasmos por permuta

Os metaplasmos por permuta ou transformação consistem na troca de um fonema por outro (Coutinho, 1976: 143). Pertencem a essa classe diversos fenômenos que ocorreram na transformação do latim para o português, como a nasalização, a sonorização, em que uma consoante surda se torna sonora (*persicum* > *pêssego*); a vocalização, em que uma consoante se torna vogal ou glide (*pectum* > *peito*), e diversos outros fenômenos. Para o estudo a seguir, vamos nos manter apenas nos metaplasmos que ocorreram durante o processo de crioulização do sãotomense.

### 6.2.1. Nasalização ou nasalização

É a conversão de um fonema oral em nasal. A nasalização das vogais geralmente ocorre devido à presença de uma consoante nasal próxima à vogal, como nos casos do português arcaico *lũa*, do latim *luna*, ou do português atual *mãe*, vindo do latim *mater*. Há casos como o de *sim*, vindo de *sic*, em que não havia consoante nasal na palavra original. Estes casos são explicados por analogias feitas pelo falante, no caso de *sim*, por influência de *non* (arc.). Também nas variantes estigmatizadas do português do Brasil ocorre tal metaplasmo, como na estrutura “eu vou vim” no lugar de “eu vou vir”. A nasalização do [vi] para [vĩ] se deve a uma questão, provavelmente, de contaminação pela forma verbal do pretérito, uma analogia, tal qual entre “*sic*” e “*non*”.

Ocorreram nasalizações em sãotomense, quase todas frutos da presença de uma consoante nasal antecedendo a vogal oral portuguesa:

- (44) medo > <sup>l</sup>mêdu  
nascer > nã<sup>l</sup>se  
aranha > a<sup>l</sup>ã  
conhecer > kô<sup>l</sup>se



A única exceção é a de vapor > vapõ (Ferraz, 1979: 37, 38), que talvez seja explicável por alguma analogia, talvez mpõ (pão), tal como se explica o “sic > sim” em português.

### 6.2.2. Palatalização

A palatalização é um metaplasmo comum na história da evolução da língua portuguesa. Consiste na transformação de fonemas consonantais alveolares [s], [z], [t] e [d] em suas contrapartes palatais [ʃ], [ʒ], [tʃ] e [dʒ]. Ocorreu em casos como: cervi[z]ia > cerve[ʒ]a, do latim para o português. Pode ser observado sincronicamente nas variantes estigmatizadas do português brasileiro, no caso das fricativas: salsicha > sal[ʃ]icha, neste caso talvez pela presença de outra consoante palatal na sílaba seguinte. Já no caso das oclusivas [t] e [d], a palatalização é um fenômeno dialetal. Ocorre no sul e sudeste do Brasil, causando mais um traço de identificação da origem geográfica do falante. No sudeste, por exemplo, diante da vogal anterior [i], a língua recua, mudando o ponto de articulação dos alvéolos para o palato, assim, no Rio de Janeiro e em alguns lugares de São Paulo em oposição ao Nordeste brasileiro, ocorrem as pronúncias: [tʃ]ia X [t]ia, e [dʒ]ia X [d]ia.

Ferraz (1979: 54-55) comenta que a palatalização de [t] e [d] ocorre no Rio de Janeiro, mas nada afirma sobre a origem desse fenômeno. Mais adiante (1979: 110) ao comentar o fenômeno no sãotomense, afirma ser essa uma influência africana, mais precisamente do quicongo. “...the origin of this type of palatalization is to be found in Kongo, where, for instance, tina “to cut” is realized phonetically as [ʃina].<sup>37</sup>” (1979: 111). Para Ferraz, houve influência africana no processo de palatalização e despatalização durante a assimilação lexical do português. É comum, durante o aprendizado de outra língua, que o falante aplique à língua aprendida regras fonéticas de sua língua materna: um falante de português brasileiro, do sudeste, ao aprender o termo inglês “teacher” [ˈtɪtʃə] pronunciar algo próximo a [ˈtʃitʃeɹ], palatalizando o primeiro [t], por apresentar-se antes de um [i]. Da mesma forma, os bakongo que inicialmente aprenderam português, poderiam aplicar as regras fonéticas do quicongo, que palataliza [s], [z], [ts], [dz], [ns] e [nz] antes de [i], (Lumwamu, 1973: 33, 34; Tim, 2008) às

---

<sup>37</sup> Ferraz (1979) grafa <c> para representar o [tʃ] do Alfabeto Fonético Internacional.

palavras portuguesas que iam sendo aprendidas, da mesma forma que o falante brasileiro palatalizaria o [t] inglês. Isso explicaria não somente a palatalização de [t] e [d], mas também de [s] e [z] antes de [i], o que é regra comum ao sãotomense.

Fenômenos distintos, mas motivados pela mesma regra, podem ter criado palavras diferentes em quicongo e sãotomense:

Empréstimo do português para o quicongo: relógio > lolonzi

Empréstimo do português para o sãotomense: relógio > lo<sup>h</sup>lozu

Em quicongo, a presença do [ʒ] antes do [i] se manteve, por ser permitida. Em sãotomense, a síncope do [i] fez com que o [ʒ] se apresentasse antes de [u], despalatalizando assim o fonema [ʒ].

Contudo, embora não se saiba se /t/ e /d/ se palatalizavam no português medieval nas mesmas situações em que ocorre tal metaplasmo no português do Brasil, sabe-se que os fonemas /tʃ/ e /dʒ/ já existiam no português seiscentista, mesmo que em situações onde hoje usamos /f/ e /z/ e não /t/ e /d/. Outra hipótese é que a palatalização pode não ter ocorrido por influência africana, mas porque o /i/, sendo uma vogal anterior, e o mais importante, uma vogal alta, pode fazer com que o recuo e subida da língua a leve ao palato, transformando, assim, /t/ e /d/ em /tʃ/ e /dʒ/, respectivamente.

### 6.2.3. Despalatalização

A despalatalização também ocorreu na história da língua portuguesa, e consiste no inverso da palatalização: o ponto de articulação deixa de ser o palato e se torna os alvéolos, [ʃ] passa a [s] e [ʒ] passa a [z]. A despalatalização também pode ser observada em variantes estigmatizadas: re[ʒ]istro > re[z]istro, e na linguagem infantil: churrasco > [s]urrasco.

A despalatalização também ocorre em sãotomense durante o processo de empréstimos lexicais do português. A seguir resumiremos as situações em que ocorreram não apenas a despalatalização, como também a palatalização.

**Tabela 23:** palatalização e despalatalização.

t, d, s, z	
Palatalizaram-se antes de /i, ĩ, j/	E mantiveram-se antes das outras vogais
tirar > tʃi'la	prato > 'platu
pedir > pi'dʒi	fundo > 'fundu
cinco > 'ʃĩku	certo > 'setu
luzir > lu'ʒi	casar > ka'za
ʃ, ʒ	
Despalatalizaram-se antes de outras vogais	E mantiveram-se antes de /i, ĩ, j/
chamar > sa'ma	peixe > 'piʃi
hoje > 'oze	fingir > fĩ 'ʒi

Ferraz (1979: 41-42, 110-111).

Como se pode observar, o fenômeno consiste em manter as consoantes palatais antes das vogais anteriores altas não-arredondadas como /i, ĩ, j/ e manter as consoantes alveolares antes das outras vogais /a/, /e/, /o/, /u/, /ɛ/, /ɔ/ e suas contrapartes nasais. Como já informado anteriormente, esta também é a regra fonética da complementização das consoantes não-palatais X consoantes palatais em quicongo (Lumwamu, 1973: 33, 34; Tim, 2008; Ferraz, 1979: 51, 52).

#### 6.2.4. Lambdacismo

Fenômeno que consiste na troca do [r] pelo [l]. Para alguns linguistas como Bagno (2003), o lambdacismo ocorre sincronicamente em português brasileiro, em falas desprestigiadas, como em: cérebro > cé[l]ebro; garfo > [galfo] > ['gawfu].

Embora não seja geral, o fenômeno ocorreu em swahili, durante o processo de empréstimos lexicais do português ocorrido ao longo da evolução da língua africana: armário > almari, tabaqueira > tabakelo (Macek, 2007).

Como não existe [r] em kongo, ao assimilar as palavras portuguesas, os primeiros falantes levaram os [r] ao [l], imprimindo o lambdacismo em praticamente todas as palavras, com raríssimas exceções. Regra geral em são-tomense, não são raros os exemplos: prato > 'platu

- (45) inteiro > 'telu  
 outro > 'otlo  
 magro > \*maglu > 'mlagu

### 6.2.5. Ieísmo

Podendo ser considerado um metaplasmo por transformação, ou ainda um *tipo sui generis* de ditongação, que por sua vez é um tipo de epêntese, o ieísmo é um metaplasmo extremamente comum em todas as línguas do mundo. Consiste na transformação do fonema [ɫ] numa semivogal [j]. É comum, sincronicamente, no espanhol. Em algumas regiões da Espanha e da América Latina, o fonema espanhol, representado grafematicamente por um “ll” é pronunciado como semivogal. Também é comum em variantes do português brasileiro faladas no interior do Brasil, na chamada linguagem caipira: palavras como filho, telhado e bolha são pronunciadas como [ˈfijju], [tejˈjadu], [ˈbojje]. A hipercorreção pode levar caipiras a pronunciarem [ˈpraɫɐ] e [aˈreɫɐ] para praia e areia. Diacronicamente, o ieísmo ocorreu também no francês. Até metade do século XVIII, palavras francesas como “billet” e “maillot”, tinham seus “ll” pronunciados como [ɫ], nas variantes prestigiadas na nobreza. Após a Revolução Francesa, com a ascensão da burguesia, a fala dos burgueses, antes estigmatizada, passou a ser prestigiada, daí a norma passou a ser pronunciar os “ll” não como [ɫ], mas como [j]. A divergência entre as pronúncias em português: bilhete (e não bi[j]ete) e maiô (e não malhô) se devem à época dos empréstimos. Bilhete entrou na língua portuguesa no início do século XVII, antes do ieísmo se tornar a norma em francês, já maiô entrou no século XX, após o processo se completar (Bagno, 2003: 33, 131).

Em sãotomense o ieísmo ocorreu com a maioria, mas não com todas as palavras. O fonema [ɫ] ainda existe, embora seja raro. Na maior parte das palavras portuguesas com [ɫ], houve o ieísmo (exemplos de Ferraz, 1979: 35).

- (46) toalha [ˈtwaɫɐ] > ˈtɔja  
agulha [ɐˈguɫɐ] > ˈguja

Em raros casos, transformou-se em [ɲ]: molhar > mɔˈɲa, ou permaneceu: orelha > ɔˈɫa, ou sofreu apócope: trabalhar > tɫabɐ.

Mendonça (1973), conhecedor do crioulo sãotomense, afirmava ser o ieísmo no português caipira uma influência africana, talvez por analogia com os crioulos, mas, por

honestidade intelectual, comentava ainda a presença do mesmo metaplasmo em espanhol e francês, línguas evidentemente sem influência africana.

### 6.3. Metaplasmos por aumento

Os metaplasmos por aumento são os que adicionam fonemas às palavras (Coutinho, 1976: 146). O acréscimo de fonemas pode acontecer no início das palavras (prótese), no meio das palavras (epêntese) ou no final das palavras (paragoge). No caso das epênteses, pode ainda ocorrer, de forma especial, para desfazer um encontro consonantal (suarabácti) como em *dogma* > *ˈdɔgime*.

#### 6.3.1. A prótese

Ao longo da evolução da língua portuguesa, o acréscimo de fonemas no início de vocábulos, ocorreu em palavras como *scribere* > *escrever*, *nanu* > *anão*, *rubore* > *arrebol*.

Nas variantes estigmatizadas do português do Brasil, ocorre em casos como *arrecife* (em vez de *recife*, tal qual o português arcaico), *amostrar* (em vez de *mostrar*), *avoar* (em vez de *voar*).

Em sãotomense ocorreram por dois motivos: a união do artigo definido português: *o pó* > *ɔˈpɔ* ou do aparecimento de uma homorgânica nasal *pão* > *mpõ* / *Galinha* [*gɛˈliɲɐ*] ou [*gɛˈlĩɐ*] > *ɲgaˈɲa*. A prótese pode ter ocorrido ainda, devido à interpretação, dos bantos, de que os artigos seriam prefixos como os da classe banta.

No caso das palavras com artigo definido, ocorreram por serem monossílabos. Em sãotomense há uma tendência para dissílabos (Ferraz, 1979):

- (47) *o mar* [*u ˈmar*] > *ɔˈmali*  
*a fé* [*ɐ ˈfɛ*] > *aˈfɛ*  
*el-rei* [*ɛlˈRej*] > *aˈle*

#### 6.3.2. A epêntese

O acréscimo de fonemas no interior de vocábulos, ao longo da história da língua portuguesa, ocorreu em palavras como: *nombro* > *numero*, *stella* > *estrela*.

No português do Brasil, nas variantes estigmatizadas, ocorre em casos como: *causo* (em vez de *caso*), e na saudação religiosa afrobrasileira ‘saravá’ (do português *salvar*, com rotacismo).

Em sãotomense, ocorreu para evitar sílabas pesadas, que são inexistentes em quicongo.

- (48) garfo > 'galufu  
maldade > mali'dadzji (Ferraz, 1979: 47).

#### 6.3.2.1. A ditongação

Caso *sui generis* de epêntese, a ditongação consiste no aparecimento de um glide (semivogal) logo após uma vogal simples. Comum na história da língua portuguesa, ocorreu para desfazer hiatos, como em *fea* > **feia**, *paes* > **pais**, *arena* > *area* > **areia**.

Quanto ao sãotomense, Ferraz (1979: 28) menciona regras sincrônicas, e não diacrônicas, das ditongações. Como não há hiatos em sãotomense, estes tendem a desaparecer, sincronicamente, por meio das ditongações.

ku a'le > kwa'le (com o rei > co'o rei)

'vĩtʃi 'lanu > vĩtʃjanu ( vinte anos > vintjanos)

Esse é o mesmo sistema existente em línguas bantas como o umbundu e o swahili (Valente, 1964: 24). Em swahili chega-se a inserir um [l] eufônico para desfazer o hiato.

#### 6.3.3. A paragoge ou epítese

Paragoge ou epítese é a adição de um fonema no final da palavra. Ocorreu durante a evolução da língua portuguesa, como em *ante* > *antes*, *preste* > *prestes*, *entonce* > *entonces* (arc.). Nesses casos, a paragoge ocorreu por analogia com a palavra depois (Coutinho, 1976: 147).

Ocorre sincronicamente no processo de empréstimos estrangeiros, como em *beef* > *bife*, *club* > *clube*, *film* > *filme*, *chic* > *chique*, nesses casos, devido às regras fonotáticas do português, que não permitem em posição de coda, nas sílabas, consoantes oclusivas (*club*, *chic*), codas complexas CC (*film*) ou a fricativa [f] (*beef*).

Em sãotomense, a paragoge ocorreu em casos como: sul > 'sulu, ar > 'ali, devido às regras fonotáticas, que não permitem um [l] em posição de coda:

- (49) lençol [lɛ'ʂɔl] > lã'ʂɔlɔ  
barril > ba'lili  
doutor > do'tolo  
cor > 'kolo (Ferraz, 1979: 47).

Ou ainda, em casos como:

- (50) dormir > dumi'ni  
funil > fu'nini  
anel > 'neni

Em que, além da paragoge de uma vogal, as consoantes em coda [l] e [r] se transformaram em [n] devido à existência de uma consoante nasal em onset na sílaba. Esses casos, são, contudo, exceções no processo fonético do sãotomense (Ferraz, 1979: 37).

#### 6.4. Metaplasmos por subtração

Os metaplasmos por subtração são os que tiram ou diminuem fonemas à palavra (Coutinho, 1976: 147). A queda de fonemas pode ocorrer no início de uma palavra (aférese), no meio de uma palavra (síncope) ou no final de uma palavra (apócope).

##### 6.4.1. A aférese

A aférese, queda de fonema em início de palavra, também ocorreu ao longo da evolução da língua portuguesa: attonitu > tonto, episcopu > bispo. Ocorre constantemente na fala não monitorada em Português: está > tá ['ta], espera aí > pera aí [pera'i], você quer > cê quer [se'kɛ].

Devido à tendência dissílaba das línguas bantas (Ferraz, 1979: 47), diversas aféreses ocorreram para transformar as trissílabas ou polissílabas portuguesas em dissílabas, caindo assim, as sílabas átonas (pretônicas ou postônicas) em português:

- (51) esquecer > ke<sup>l</sup>se  
embrulhar > bu<sup>l</sup>ja

#### 6.4.2. A síncope

A supressão de um fonema no interior da palavra, a síncope, ocorreu ao longo da evolução da língua portuguesa em casos como *mediu* > *meio*, *malu* > *mau*, *manica* > *manga*, *opera* > *obra*.

Podemos observar casos de síncope em variantes estigmatizadas do português do Brasil, como: *abóbora* > *abobra*, *xícara* > *xicra*, e, contraditoriamente, em variantes prestigiadas, *aboborinha* > *abobrinha*, *xicarinha* > *xicrinha*, *para* > *pra*, *barezinhos* > *barzinhos*, *colherezinhas* > *colherinhas*<sup>38</sup>.

A mesma regra que gerou as aféreses em *sãotomense*, as tendências ao dissilabismo, levou às síncopes:

- (52) *xícara* > <sup>l</sup>fíkla  
*barato* > <sup>l</sup>blatu  
*máscara* > <sup>l</sup>mlafka

#### 6.4.2.1. A monotongação

Tipo *sui generis* de síncope, em que um glide desaparece, a monotongação é um fenômeno comum na língua portuguesa (cf. seções 4.2.1.5. O ditongo [gũ], 4.2.1.6. O ditongo [ow], 4.2.1.7. O ditongo [ej]).

No exemplo a seguir, do *sãotomense*, houve uma monotongação de [aj] para [a]. O caso de [õ] na verdade não foi monotongação, já que a palavra original, no século XVI, era provavelmente *caixom*.

- (53) *caixão* [kaj<sup>l</sup>fõ] > kasõ  
cf. o swahili, língua sem ditongos, *caixa* > *kafa* (Macek, 2007).

As monotongações são bastantes comuns em *sãotomense*. Ditongos como [ej] e [ow] desaparecem por completo em qualquer posição.

---

<sup>38</sup> Há de se ressaltar que o caso de *barzinhos* e *colherinhas* talvez seja morfológico e não fonético.



- (54) pedreiro > pedlelu  
peito > petu  
(cf. seção 4.2.1.7. O ditongo [ej])

Uma explicação possível para tal fenômeno é a ausência de ditongos na língua de substrato, o quicongo (Coene, 1960: 1), conforme foi observado na seção 3.

#### 6.4.3. A apócope

Apócope é a queda dos fonemas no fim das palavras. Ocorreu, na língua portuguesa, em casos como amare > amar, amat > ama, regale > real. Ocorre constantemente em todas as variantes do português brasileiro, em todos os níveis socioculturais, com o ‘r’ em posição de final de palavra, sobretudo nos verbos no infinitivo: amar > [a<sup>h</sup>ma], querer isso > [ke<sup>h</sup>re <sup>h</sup>isu], mas não em casos como ‘por isso’.

Em sãotomense, ainda para fazer os dissílabos, ocorreram algumas apócopies, como em trabalhar [traba<sup>h</sup>lar] > tla<sup>h</sup>ba, não sem que antes houvesse uma sístole, obviamente.

#### 6.4.4. A sinalefa ou elisão

Sinalefa é a queda de vogal final de uma palavra quando a seguinte começa por vogal. Ocorreu em casos como: de + intro > dentro, de + aquele > daquele. Ocorre constantemente na fala não monitorada em português, qualquer que seja a variante. Para + o > pro, de + uma > duma, caixa de água > caixa d’água.

Em sãotomense, ocorre sincronicamente (Ferraz, 1979: 28).

- (55) <sup>h</sup>dudu di <sup>h</sup>awa > <sup>h</sup>dudu dawa (jarro de água > jarro d’água)

Trata-se, pelo visto, de um fenômeno fonético comum, sem influência africana, já que a forma sem elisão respeita todas as regras fonotáticas, assim como as formas com elisão.

#### 6.5. Metaplasmos por transposição

Os metaplasmos por transposição consistem no deslocamento de um fonema ou de um acento tônico da palavra (Coutinho, 1976: 148).

### 6.5.1. A metátese

A metátese é a transposição de fonemas, que se pode verificar na mesma sílaba ou entre sílabas. Ocorreu durante a evolução do latim para o português: *semper* > *sempre*, *pigrítia* > *pegriça* (arc.) > *preguiça*, *rabie* > *ravia* (arc.) > *raiva*. Sincronicamente, ocorre também nas variantes estigmatizadas: *tábua* > *tauba*, *estupro* > *estrupe*, *lagartixa* > *largatixa*.

Foram extremamente comuns em sãotomense devido à estrutura silábica. Em sãotomense, um 'l' jamais pode permanecer em posição de coda. Devido a essa regra, muitas metáteses ocorreram para formar sílabas CVC em CCV, com o 'l' na posição de C<sub>2</sub>:

- (56) *lagartixa* > *laga<sup>l</sup>tlixa*  
*jardim* > *ʒa<sup>l</sup>dĩ*

Outros casos como *vidro* [<sup>l</sup>vidru] > <sup>l</sup>vlidu, não devem ter ocorrido por regras fonotáticas, mas por simples deriva, já que <sup>l</sup>vidlu é gramaticalmente perfeita.

### 6.5.2. hiperbibasmo

Consiste na alteração do acento tônico para uma sílaba anterior (sístole) ou posterior (diástole). O hiperbibasmo ocorreu ao longo da evolução da língua portuguesa: *amavissémus* > *amassémus* > *amássemos* (sístole), *océano* > *oceanos* (diástole). (Coutinho, 1976: 149).

Em sãotomense o hiperbibasmo ocorre por flutuação livre (Ferraz, 1979: 107). Em línguas não crioulas não há tanta flutuação fonológica como nos crioulos. Não se trata de variação dialetal, visto que, para Ferraz (1979: 8) não existe tal variação nos crioulos do Golfo da Guiné. Trata-se de variações feitas pelo próprio falante: *nisidadzi* > *misi<sup>l</sup>dadzi* (Ferraz, 1979: 108) em que um mesmo falante varia a pronúncia das palavras sem razão específica. Tal flutuação pode ocorrer com vogais: *ɔmɛ* ~ *ɔme*, com consoantes: *ska* ~ *ʃka*, nasalidade: *mu* ~ *mũ*, e finalmente, a sílaba tônica, *ʔzanta* ~ *zan<sup>l</sup>ta*, *a<sup>l</sup>le* ~ *ʔale*. (Ferraz, 1979: 108).

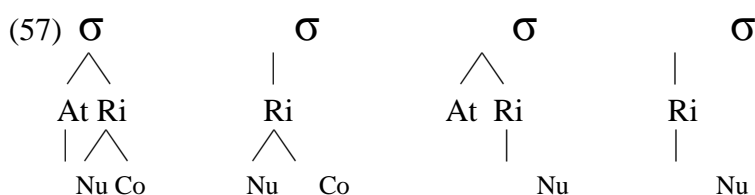
## 7. OUTRAS CONTRIBUIÇÕES DO SUBSTRATO

Nesta seção tentaremos mostrar outras contribuições que as línguas de substrato deram aos crioulos do Golfo da Guiné, além das já apresentadas neste trabalho. Embora centrado no sãotomense, mostraremos, quando for pertinente, as contribuições também em outros crioulos da mesma área, como o principense, o angolar e o anobonês.

Algumas das contribuições no campo da fonologia são a estrutura silábica e o desaparecimento do flepe [ɾ] e do trill [r] no sãotomense; na morfologia notamos um sistema de plural diferente do português e das línguas bantas, e mais próximo das línguas edoides; na sintaxe, observamos os pronomes disjuntivos em locuções prepositivas e a negação disjuntiva; no léxico, mostraremos as contribuições africanas em termos da lista-swadesh, e por fim, os ideofones, elementos linguísticos tão comuns nas línguas africanas, embora não restritos a elas.

### 7.1. Estrutura silábica do sãotomense e do principense em comparação com as estruturas silábicas das línguas negro-africanas

De acordo com Creissels (1994), há apenas 4 possibilidades teóricas nas realizações de sílabas em línguas negro-africanas, a saber:



Sendo  $\sigma$  a sílaba; **At** o ataque (fonema que abre a sílaba); **Ri** a rima, parte da sílaba que se inicia com a vogal tônica ou subtônica e, dentro da estrutura da rima: **Nu** o núcleo (vogal) e **Co** a coda (fonema que “fecha” a sílaba), podendo ser uma vogal – nas sílabas leves – ou consoante – nas sílabas pesadas.

As sílabas pesadas são raras em línguas negro-africanas (Creissels, 1994). Em português europeu as sílabas pesadas com oclusivas em posição de coda são comuns (*dogma*, *advogado*, *pacto*, *ritmo*), mas desapareceram tanto no português do Brasil quanto nos crioulos de São Tomé e Príncipe, mas não no português europeu.

Nas línguas negro-africanas, quando ocorrem sílabas pesadas, geralmente a coda é uma nasal. Mesmo nas línguas africanas que possuem sílabas pesadas (minorias), estas não seguem o esquema comum de sílabas pesadas de línguas como o inglês. Em wolof (Senegal), /p, c, k/ em coda são realizados como não-explosivos. Em zarma<sup>39</sup> as sílabas fechadas ocorrem com consoantes geminadas. (Creissels, 1994).

A estrutura silábica das línguas africanas do substrato, principalmente o quicongo que só possui sílabas CV (Coene, 1960: 1; Lumwamu, 1973: 25; Ferraz, 1979) pode ter servido de base para a formação das sílabas em crioulos, adaptando, à nova realidade, todo vocábulo que veio do português.

The syllable structure of ST is typically CV, as **is characteristic of Bantu and as is found in some Kwa<sup>40</sup> languages, such as Bini**. This provides a contrast with Creoles with a **different substratum** which may have closed syllables, as may be exemplified by the Creole of the Cape Verde Islands, which has a substratum which is predominantly West Atlantic. (...) Portuguese closed syllables have been made open in ST, but have remained closed in Creole of Cape Verde (...)

Ferraz (1979: 111), grifo nosso.

A comparação que Ferraz faz entre a estrutura silábica do sãotomense e do caboverdiano é pertinente. Se a estrutura silábica de um crioulo de base portuguesa se aproxima mais de seu substrato africano enquanto a estrutura de outro crioulo permanece como no português europeu, porque atribuir isso a simples evolução da língua quando sabemos que no segundo caso há proximidade estrutural com o substrato africano? O problema das estruturas silábicas nos leva, sem dúvida, à problemática comparação entre o português europeu e do português do Brasil. Enquanto o primeiro possui um amplo inventário de sílabas pesadas tanto em onset quanto da coda, o segundo as limita, como se demonstra nas suarabáctis que comumente se fazem no português falado para vocábulos como pneu [pi'nɛw] ~ [pe'nɛw], advogado [adʒivo'gadu] ~ [adevo'gadu], dogma ['dɔgimɛ], ritmo ['χitimu], dentre outras. A tendência do português brasileiro para esse fenômeno já pode ser observada no século XIX, como se pode ver em versos de poetas do romantismo brasileiro, que parecem se aproveitar de uma possível variação fonética para facilitar a metrificação das sílabas. Em I-Juca-Pirama, de Gonçalves Dias, pode-se ver a suarabácti no decassílabo: “Contudo os olhos d’ignóbil pranto” (canto II, verso 13) em que a suarabácti de ‘ignóbil’ favorece o surgimento de uma sílaba a mais, sem a qual o verso se quebraria

<sup>39</sup> Língua do tronco nilo-sahariano falada no Níger, na Nigéria e em Burkina Faso.

<sup>40</sup> Em 1979, kwa; atualmente, edoide (cf. Gordon, 2005).

com os outros decassílabos do canto. A tendência do português do Brasil a ter sílabas leves pode vir de uma evolução natural da língua, bem como de uma influência fonética africana ou indígena. Sabe-se, por levantamento de vocabulário, que o português do Brasil recebeu influências lexicais do quimbundo e do iorubá<sup>41</sup>. Em quimbundo, tal qual na maior parte das línguas africanas: “Todas as sílabas terminam em vogal e nunca em consoante” (Quintão, 1934: 12).

## 7.2. Influência da estrutura silábica africana no principense

Enquanto o são-tomense conta com sete tipos de sílabas (cf. seção 6.5), sendo 88% das sílabas do são-tomense V ou CV, o principense conta com um inventário silábico mais reduzido, cinco tipos de sílabas, sendo que 95% das sílabas do principense são V ou CV (cf. seção 10.7).

(58)	<b>Português</b>	<b>são-tomense</b>	<b>principense</b>
	Intestino (tripa)	[ˈtɫipa]~[ˈqɫipa]	[ˈtʃipa]
	Coração	[ˈqɫɔsõ]~[ˈkɫɔsõ]	[ˈkosã]
	Cabra	[ˈkabɫa]	[ˈkaβa]
	Cobra	[ˈkɔβɫo]	[ˈkobo]

(Dados retirados de Graham & Graham, 2004).

Pode-se observar que, enquanto o são-tomense apresenta clusters complexos como /tɫ/ em /tɫipa/, o principense dá preferência à consoante [tʃ] como em [tʃipa], da mesma forma, o /kɫ/ do são-tomense /kɫɔsõ/ se apresenta em principense como /k/ em /kosã/. O principense tem uma tendência maior às sílabas CV que são 95% do que o são-tomense que possui 88% de sílabas dessa natureza em seu léxico.

O sistema silábico do principense só não apresenta uma ocorrência de 100% devido à influência que a língua ainda sofre do português, segundo Günther (1973: 45):

<sup>41</sup> Quimbundo é uma língua banta falada em Angola e da qual vieram palavras como carimbo, maconha e moleque. Iorubá é uma língua ioruboide (antigamente kwa) falada na Nigéria, no Benin e no Togo, também conhecida como jeje e língua-do-santo, e de onde vieram palavras como acarajé e orixá.

Ursprünglich kannte das Principensische nur offene Silben, d.h. jede Silbe ging auf Vokal aus. Diese Auffassung wird bestätigt durch die Lautgesetze (vergl. Appendix III). Da aber das Portugiesische einen ständigen Superstratseinfluß hatte und noch hat, kommen auch Beispiele mit Konsonantanhäufungen strukturell gesehen: geschlossenen Silben (...)<sup>42</sup>

No mencionado apêndice III, Günther (1973: 266) apresenta *Die wichtigsten Lautwandel von Portugiesischen zum Principensichen* (Importantes transformações do português para o principense). Nesta parte de seu livro, demonstra como algumas consoantes portuguesas permaneceram em principense e como outras sofreram metaplasmos, além de algumas alterações ocorridas na estrutura da sílaba. Estas alterações de estrutura silábica aparecem na presença de /n, l, χ, r, s, ʃ, ʒ, z/. ou seja, exatamente as consoantes que podem ocupar posição de coda no português do Brasil (já no português europeu outras consoantes podem ocupar essa posição).

Assim Günther (1973: 267), a título de exemplo, descreve algumas transformações fonológicas, como quando um /r/ que ocupa uma posição entre C e V ou entre V e C em português, desaparece em principense.

(59) preto > p<sup>h</sup>étu – síncope do [r] – CCV.CV > CV.CV

Ou ainda:

(60) arroz > oróso – paragoge – V.CVC > V.CV.CV

amarrar > mará – aférese do [a] e apócope do [r] – V.CV.CVC > CV.CV

verde > véde - síncope do [r] – CVC.CV > CV.CV

A queda do [r] em núcleo de "preto" e em coda de "verde" e "amarrar" serviu principalmente para criar sílabas CV, aproximando assim o esquema silábico do principense e de línguas africanas. Como o principense possui uma influência bini maior que quicongo, apresentam-se em principense mais sílabas V (14%) do que em são-tomense (7%). Em bini as sílabas V são comuns, enquanto em quicongo não existem.

---

<sup>42</sup> Inicialmente conheciam-se no principense apenas sílabas abertas, isto é, cada sílaba terminava por vogal. Esta opinião se confirma por meio das “regras fonológicas” (Lautgesetze) (vide Apêndice III) (...) Contudo, o português exerceu influência contínua no principense e ainda a exerce, por isso, temos exemplos também de palavras (...) com sílabas fechadas. (Tradução nossa)

### 7.3. Influências fonológicas africanas no sãotomense

#### 7.3.1. O desaparecimento da vibrante múltipla /r/ e do tepe /ɾ/

Em sãotomense não existe a consoante vibrante múltipla do português seiscentista /r/ (trill), e o tepe /ɾ/ é raro. Em quicongo não existem os fonemas /r/ e /ɾ/. O /l/ é a única consoante lateral existente em quicongo. O fato de em sãotomense também não existir /r/ e haver um raro uso para o tepe /ɾ/ se explica pela influência fonológica do substrato africano. Segundo Ferraz: “This has a basis in Kongo dialects as Kishikongo, where /l/ is the only liquid which occurs.” Para comprovação, o autor cita exemplos de palavras que entraram tanto no sãotomense quanto no quicongo, nas duas línguas com lambdacismo (Ferraz, 1979: 112):

(61)	<b>Português</b>	<b>sãotomense</b>	<b>quicongo</b>
	hora [ˈɔrɐ]	ˈɔla	ˈɔla
	relógio [rɐˈlɔʒju]	ˈlɔlɔzu	loˈlonʒi

Quando da entrada do léxico português num pidgin de falantes de uma língua sem esses dois fonemas, os falantes assimilaram o léxico com algumas características. Quando o tepe ou trill (vibrante múltipla) está em posição de ataque ou núcleo, ocorre lambdacismo, quando estão em coda absoluta, há queda, tal como em algumas variantes do português do Brasil nos finais de infinitivo: “amar” [aˈma] e substantivos como “lugar” [luˈga].

**Tabela 24:** a transformação do [r] e do [ɾ].

Posição	Português	Sãotomense	Fenômeno
Ataque absoluto	rato [ˈratu]	ˈlatu	Lambdacismo
	relógio [rəˈlɔʒju]	lɔˈlɔzu	
Ataque simples	ferro [ˈfɛ.ru]	ˈfɛlu	Lambdacismo
	tirar [tiˈrar]	tʃiˈla	
Núcleo	outro [ˈo.tɾu]	ˈotlo	Lambdacismo
	grande [ˈgrã.di]	ˈnglãdʒi	
Coda	largo [ˈlar.gu]	ˈlalugu	Lamb+epêntese
	serviço [sər.ˈvi.su]	ʃliˈvisu	
	inferno [iˈfɛr.nu]	ˈfɛnu	Queda do R
	Verde [ˈvɛr.di]	ˈvede	
Coda absoluta	Descer [dɛʃˈsɛr]	deˈse	Queda do R
	Tirar [tiˈrar]	tʃiˈla	

Como o português continua exercendo influência sobre o sãotomense e como grande parte dos falantes são bilíngues ou trilíngues, devido à língua oficial ser a portuguesa, vem ocorrendo cada vez mais um rotacismo, nas variantes urbanas e escolarizadas (Araújo, 2007). Inicialmente o tepe funcionava como alofone do /l/. O /r/ também pode ser encontrado, mas é extremamente raro. O processo de rotacismo já criou pares mínimos no sãotomense:

(62)	<b>Português</b>	<b>sãotomense</b>
	Caldo [ˈkaldu]	ˈkalu
	Caro [ˈkaru]	ˈkaru
	Carro [ˈkaru]	ˈkaru

(Ferraz, 1979: 37).

Apesar de [ɾ] e [r], por causa desses pares mínimos, poderem ser considerados fonemas, como o faz Mane (2003), o uso de tepes e trills é raro e restrito como afirma Ferraz (1979: 37), e se deve a um processo de descrioulização, não pertencendo o /ɾ/ ou o /r/ tradicionalmente ao inventário fonológico do sãotomense.



No verbete enciclopédico online sobre o quicongo, escrito no próprio idioma quicongo (Wikipedia, 2007)<sup>43</sup>, há uma listagem de palavras portuguesas que entraram no idioma:

Bangogo ya kimputulukesi:<sup>44</sup>

(63)	kilapi	( <i>lápiz</i> )
	kumbi	( <i>comboio</i> )
	mapa	( <i>mapa</i> )
	misa	( <i>missa</i> )
	Mputu	( <i>Portugal</i> )
	numelo	( <i>número</i> )
	sikoti	( <i>chicotte</i> )
	wanzio	( <i>anjo</i> )

Pode-se perceber nessa curta lista de palavras alguns fenômenos fonológicos já observados no crioulo sãotomense, a saber:

Lambdacismo: número > numelo, como é regra geral no sãotomense. Epêntese, anjo > wanzio [wanzio]. Como a consoante [ʒ], pelas regras do quicongo, não pode aparecer antes de [o] ou [u] e somente de [i], uma vogal epentética surgiu para manter o fone [ʒ]. Outra possibilidade seria a despatalização [ʒ] para [z] como ocorreu em empréstimos franceses: jamais > zamé.

#### 7.4. Influências sintáticas

##### 7.4.1. Pronomes disjuntivos

Uma das influências africanas da sintaxe no sãotomense está na presença de pronomes disjuntivos em locuções prepositivas, inexistentes em português mas comuns em línguas africanas (Ferraz, 1976).

(64) **'nadzi sũ bi de?**

Onde 2SG vir d-3PL?

De onde o senhor vem? (Ferraz, 1976).

---

<sup>43</sup> A fonte não é confiável, mas foi mantida por não ter sido encontrada a mesma informação em outra fonte.

<sup>44</sup> “Palavras [oriundas] do português”

- (65) **l'ãdʒi bo ka ta 'n-e?**  
 onde 2SG KA estar em 3PL?  
 Onde você mora? (Ferraz, 1979: 64).

#### 7.4.2. Negação disjuntiva

Conforme já mencionado no item 7.13, a negação disjuntiva consiste na utilização de duas partículas para negação, uma antes do elemento negado e outra no final da sentença. Em sãotomense a negação se faz com as partículas 'na' e 'fa'.

- (66) **l'inē na kō'se 'piʃi 'godo fa**  
 2PP NEG<sub>1</sub> conhecer peixe gordo NEG<sub>2</sub>.  
 Eles não sabem onde tem peixe bom. (Ferraz, 1979: 65).

A negação disjuntiva ocorre não apenas em sãotomense (na...fa), mas também em angular (na...wa) e anobonês (na...f) (Ferraz, 1979: 10). Em principense usa-se apenas a partícula (fa) ao final da frase (Günther, 1973: 51; Ferraz, 1979: 10). A mesma estrutura de negação disjuntiva pode ser encontrada em diversas línguas bantas, como em bangi, ngola (Ferraz, 1979: 11), quicongo (Dereu, 1955), lembaama (Okodowa, 2005). Dereu (1955: 29) apresenta frases em quicongo em que aparecem construções com dupla negativa que, segundo o autor, podem ser ke...ko ou ka...ko:

- (67) Ngi-na  
 1SG-ser  
 Eu sou
- (68) Wa mbote ngi-na  
 PREF<sup>45</sup> bom 1SG-ser  
 Eu sou bom

---

<sup>45</sup> "Wa" é prefixo de primeira classe, faz o plural em "ba". Na conjugação verbal mostrada por Dereu (1955: 29) aparece nas três pessoas do singular, e ba aparece nas três pessoas do plural. Note a mesma frase com o pronome no plural, em que na posição de wa aparece ba.

- (69) k'ina            wa    mbote ko  
 NEG<sub>1</sub>-1SG-ser PREF bom    NEG<sub>2</sub>  
 Eu não sou bom
- (70) Tu-na  
 1PL-ser  
 Nós somos
- (71) ba    mbote tuna  
 PREF bom    1PL-ser  
 Nós somos bons
- (72) ka    tuna    ba    mbote ko  
 NEG<sub>1</sub> 1PL-ser PREF bom    NEG<sub>2</sub>  
 Nós não somos bons.

A origem precisamente da partícula “fa” não é conhecida. Ferraz (1979: 11) informa que em lolo<sup>46</sup> existe uma partícula “fa”, negativa, contudo é usada prefixalmente, sem disjunção alguma.

#### 7.5. Influências lexicais e semânticas do quicongo e do bini no sãotomense

As línguas africanas exerceram forte influência não só no léxico do sãotomense, como também no modelo semântico de palavras vindas do português. Em línguas africanas, uma única palavra serve para designar as palavras pé e perna, e outra única para “mão” e “braço”. Desta forma, a palavra “pé” entrou no sãotomense (o'pe) com o significado de “pé” e “perna”, e a palavra mão (mõ) com o significado de mão e braço, deixando de fora do sãotomense palavras portuguesas como perna e braço, que de nada serviriam num léxico com extensão de modelos semânticos africanos. A palavra “blasu” existe em sãotomense, mas significa abraço, e vem desta última, por aférese e lambdacismo. Myers-Scotton chama de convergência quando parte da estrutura abastrata vem de outra língua, mesmo que o léxico seja português.

---

<sup>46</sup> Língua banta falada por 15 mil pessoas em Moçambique, no oeste da província de Zambézia (centro do país).

Já foi dito anteriormente que 93% do corpus lexical do sãotomense (cerca de 1.400 palavras), levantado por Negreiros (1895) e citado por Ferraz (1979) é de origem portuguesa. Os 7% restantes, palavras de línguas africanas, compõe-se de 96 vocábulos levantados, sendo 60 de origem banta e 36 de origem kwa <edoide> (Ferraz, 1979: 114-115).

No português atual, segundo uma simples pesquisa realizada no dicionário eletrônico Aurélio (de 435 mil verbetes), listaram-se 279 verbetes com etimologia do quimbundo e 222 com etimologia do iorubá, que somadas dão pouco mais que 0.1% do vocabulário total. Numericamente, a influência africana no léxico do português do Brasil se mostra mais fraca do que a de outras línguas. Para comparação, o mesmo dicionário traz 2.125 palavras portuguesas vindas do inglês, 2.533 vindas do tupi, 3.436 vindas do francês e 6.767 do árabe.

Analisando pormenorizadamente as cerca de 500 palavras de origem africana listadas, a maior parte se apresenta em contextos bastante peculiares, como o campo religioso (abadô, babalaô, xeré, do iorubá), da culinária (abará, acarajé, ebó do iorubá) ou de termos referentes à escravidão (senzala, mocambo, mucama, banza do quimbundo<sup>47</sup>). Fora as palavras de origem africana que só são usadas no português falado em África, já que o dicionário levanta o léxico do português mundial, e não apenas brasileiro. O mesmo dicionário traz ainda mais de 100 palavras cuja origem etimológica é do sãotomense (forro), mas pelo que se nota dos exemplos dados no dicionário, embora possam ser termos correntes no português cotidiano de São Tomé e Príncipe, parecem mais de uso literário, como “flogá” que entrou no dicionário por ter sido registrada na obra "Tempo de Flogá", do escritor africano Sum Marky (Dicionário Aurélio, 2006: verbete “flogá”). Essa palavra, vinda do português europeu do século XVI “folgar” com sentido de divertir-se, fez o caminho inverso e voltou para o português, desta vez o português literário atual de São Tomé e Príncipe.

Observando a lista-swadesh do sãotomense, observamos influências lexicais africanas, como em partes do corpo (ʼbezu-beʼzu - queixo, kokoʼi - cotovelo, pimʼbi - pênis, etc.).

---

<sup>47</sup> Não negamos a influência do quimbundo no léxico de uso cotidiano brasileiro (cochilar, moleque, cafuné, carimbo, cachimbo, xingar, samba, caçula, etc.) contudo, na lista-swadesh do português do Brasil, há apenas uma palavra: “bunda”, quando em sãotomense o número é bem maior.

## 7.6. Os ideofones

A mais evidente das influências africanas nos crioulos do Golfo da Guiné é o sistema de ideofones.

O termo ideofone tornou-se comum no início das descrições de línguas indígenas, e já foram dadas muitas definições para o termo, muitas das quais confundindo os ideofones com onomatopeias, ou dando a eles classes morfológicas distintas dos verbos ou substantivos. Essas confusões são naturais em qualquer ciência sempre que algo está em fase de estudo inicial, e são corrigidas com o passar do tempo pelos pesquisadores posteriores. Por esse motivo, ficaremos com a definição de Dumestre (1998), que surgiu de estudos aprofundados dos ideofones em Bambará, e definições feitas anteriormente por outros pesquisadores. Para Dumestre (1998: 322), ideofones são todo elemento da língua que associa um fonetismo e um semantismo particulares. Por fonetismo particular, entende-se como características fonéticas estatisticamente marginais: consoantes não fonológicas ou de uso emprego raro na língua, como o [z] em início de palavras no francês, conforme Dumestre (1988) citou sobre sua própria língua) palavras com um número expressivo de sílabas, como o termo “muito branco” em anobonês, que se diz “banco pepepepe”, etc. Por semantismo particular, entendem-se as associações à expressividade, emoção, afetividade, por meio de certas onomatopeias, termos engraçados, infantis, gírias, e geralmente de elementos mais designativos que evocativos. Para Dumestre, os ideofones, segundo sua definição, são atestados em todas as línguas. Os ideofones, apesar de existirem em todas as línguas, são mais comuns em línguas africanas e ameríndias que nas línguas do tronco indo-europeu. No português, por exemplo, são raros os contextos em que aparecem, e quando aparecem, são formados quase sempre por reduplicação. Alguns exemplos de ideofones no português:

1. Na linguagem infantil: bebeça (cabeça), bumbum (do quimbundo), dodói (ferida), dudu, mimi (dormir), nenem, papato, pepé (pé); alguns evidentemente de origem onomatopaica: auau (cão), bibi (carro), cocó (galinha), pipi, xixi (origem obscura, provavelmente onomatopaica).

2. Em apelidos antroponímicos: Dedé (de André), Dudu (de Eduardo), Zezé (de José ou Maria José), etc.

3. Em expressões com o significado de conversa vazia: nhenhém (do tupi *nhe'eng* – falar), blabláblá, lenga-lenga, papapá, tralalá, trololó, etc.

4. Reduplicação interna: pititinha (muito pequeno).

5. Reduplicação tradicional portuguesa: vovô, vovó, papai, mamãe, titio(a), nenê, bebê, babá, reco-reco, corre-corre, fuque-fuque, bole-bole, come-come, balaco-baco, mata-mata, totó (pebolim), cricri, etc.

Por serem raros em português, e mais comuns na linguagem infantil, os ideofones chegam a criar problemas de tradução, quando se vertem textos de línguas repletas de ideofones para línguas em que são raros, como enfatizou Crofts (2006) em relação aos ideofones da língua indígena brasileira munduruku:

Nas narrativas em mundurukú, é muito frequente o uso de ideofones. O papel deles é tornar as narrativas mais animadas. Em uma narrativa bem viva ocorrem no mesmo parágrafo entre cinco e vinte ideofones. Alguns parágrafos consistem mais em ideofones do que em sentenças completas. Contudo, **quando estes ideofones são traduzidos literalmente para o português, a tradução assim feita pode dar a impressão de ser uma linguagem infantil.** Na realidade, não é este o fato. A fim de evitar que haja esta interpretação errada, recomenda-se que a tradução em português omita os ideofones, ou que sejam empregados recursos da língua portuguesa, obtendo-se assim o resultado narrativo equivalente aos ideofones em mundurukú. (grifo nosso)

Em seu livro sobre a influência das línguas africanas no português do Brasil, Mendonça (1973: 69) nada fala sobre ideofones, pelo menos não diretamente, nem usando tal terminologia, contudo, quando menciona as influências morfológicas, diz: “A própria linguagem infantil tem um sabor quase africano: cacá, pipi, bumbum, nenem, tatá, papato, lili, mimi, cocô, dindinho, bimbinha.”

O desconhecimento dos ideofones e seu largo uso em línguas africanas chegou a criar, no início de seu estudo, descrições que demonstravam determinado espanto dos gramáticos, como se vê em Barrena (1957), em sua descrição do anobônês, o que demonstra serem os ideofones realmente raros em línguas europeias:

Para designar el más alto grado del aumentativo o diminutivo usan una **forma especialísima y muy curiosa**, según lo indican los varios ejemplos que se ponen a continuación, como superlativo.  
Muy seco – sécu jolójólo  
Muy mojado – madu potopóto  
Muy dulce – menémenéméne  
Muy blanco – bancu pepepepe.  
(grifo nosso)

Assim como ocorre em anobônês; em sãotomense, os ideofones são tão comuns quanto nas línguas africanas que lhe serviram de substrato. São, em sua totalidade, retirados das mesmas.

- (73) (vɛ) **'kɛte-ke'te** muito velho  
 (lu'zi) **nge'ne-nge'ne** brilhar fortemente

(Ferraz, 1979: 98).

Os ideofones em sãotomense podem modificar um substantivo:

- (74) **Nõ 'mese nge sasa'sa**  
 1PL querer pessoa IDEOFONE  
 Queremos uma pessoa animada.

Um adjetivo:

- (75) **ka'fe sa 'kētji zuzu'zu**  
 café ser quente IDEOFONE<sup>48</sup>  
 O café está muito quente.

Um verbo:

- (76) **'kɔpu ke'bla 'nwini'nwini**  
 copo quebrar IDEOFONE  
 O copo quebrou em pequenos pedaços.

A forte presença de ideofones em crioulos demonstra mais uma influência clara do substrato africano. A presença de ideofones em português pode-se dever a uma tendência universal das línguas.

---

<sup>48</sup> Zuzuzu é o ideofone. Seu significado é “muito”, conforme definição do dicionário online de sãotomense <<http://pt.wiktionary.org/wiki/Categoria:S%C3%A3o-tomense>>.

## 8. OUTRAS CONTRIBUIÇÕES DO SUPERSTRATO

### 8.1. Fonologia

O sistema fonológico do sãotomense não é um decalque do sistema fonológico quicongo ou bini, tampouco um decalque do sistema português, mas um sistema que une características de ambos. A presença de fonemas inexistentes nas línguas de substrato como [ɾ], [ɲ] e [ʎ], a ausência de tons, a palatalização de [s] e [z] em coda e a existência de sílabas complexas como CCV e CCVC demonstram que a influência da língua portuguesa não ficou apenas no léxico, aprofundando-se até a fonologia, como já foi demonstrado ao longo deste trabalho.

### 8.2. Morfologia

O sistema morfológico banto é baseado num sistema de classes morfossemânticas, em que o singular e o plural dos substantivos são formados pela troca de prefixos, como no quicongo: muntu (homem), bantu (homens), enquanto em português é usado um sistema de desinências e afixos herdados do latim (homem, homen-s).

Uma das influências da morfologia portuguesa no crioulo de São Tomé é a formação dos participípios. Em português os participípios são formados com (Radical + vogal temática + d + desinência de gênero), como em amar – amado – amada, vender – vendido – vendida, partir – partido – partida. Como em sãotomense não são comuns as desinências de gênero, como será demonstrado mais adiante, os participípios são formados com “-du” acrescentado ao verbo, independentemente da origem etimológica deste:

(77)	<b>Forma verbal</b>	<b>Significado</b>	<b>Particípio</b>	<b>Etimologia</b>
	ma <sup>l</sup> la	amarrar	ma <sup>l</sup> ladu	do português “amarrado”
	ta <sup>l</sup> sõ	sentar	ta <sup>l</sup> sodu	do português “estar no chão”
	fo <sup>l</sup> no	chorar	fo <sup>l</sup> nodu	do kwa fo <sup>l</sup> no

Alguns substantivos possuem duas formas, uma para o masculino e outra para o feminino, como — <sup>l</sup>sɔglu / <sup>l</sup>sɔgla — sogro / sogra, além, claro, dos já citados homem /



mulher. Trata-se de uma influência portuguesa, já que em quicongo e bini não há gênero. Apesar da existência de substantivos com duas formas, há também em são-tomense substantivos uniformes, que fazem o feminino com o acréscimo da palavra 'mwala —fêmea—, do português “mulher”, e a ênfase do masculino é formada com o acréscimo de 'ɔmɛ —macho—, do português “homem”.

- (78) bwe 'ɔmɛ      boi macho (boi)  
      bwe 'mwala    boi fêmea (vaca)

Pode parecer desnecessário o acréscimo do elemento 'ɔmɛ após os substantivos para indicar que são do gênero masculino, já que o substantivo em sua forma primitiva já indica masculino singular. O uso desse termo pode, apesar disso, servir para enfatizar o gênero ou sexo do ente nomeado.

No português do Brasil pode-se usar os substantivos homem e mulher para enfatizar o sexo já mencionado na desinência do gênero colada ao radical, como nos casos: menino homem / menina mulher, filho homem / filha mulher, médico homem / médica mulher e outras centenas de casos.

## 9. CONTRIBUIÇÕES CONJUNTAS E ELEMENTOS AUTÊNTICOS

Além das contribuições do substrato e do superstrato facilmente detectáveis, o sistema gramatical do são-tomense apresenta características que coincidem com a língua de superstrato e as línguas de substrato, tornando difícil perceber se são contribuições do substrato, do superstrato ou se são contribuições conjuntas, isto é, se as línguas de substrato e superstrato contribuíram juntas para as devidas características do são-tomense, como, por exemplo, a ordem substantivo-adjetivo, a mesma existente em português e quicongo. Se a ordem substantivo-pronome demonstrativo é usada em são-tomense, tal qual em quicongo e inversamente à ordem portuguesa, e nessa característica podemos rastrear a influência africana, o mesmo não ocorreria se essa fosse também a ordem portuguesa.

Ferraz (1979: 60-61) dedica um curto capítulo com menos de uma página para discorrer sobre a estrutura dos nomes em são-tomense, e, apesar de mencionar “Forms of nouns and adjectives” nada fala sobre os adjetivos nessa seção. Devido a isso, é de se deduzir que os adjetivos se comportem exatamente como os substantivos, apresentando suas formas primitivas no masculino e no singular.

Mais informações sobre os adjetivos são dados na página 80, onde Ferraz informa-nos sobre a posição dos adjetivos em relação aos substantivos. Ao contrário da posição da língua inglesa, que Ferraz utilizou para escrever seu livro, os adjetivos são colocados pospostos aos substantivos, assim como em português, kikongo e bini.

(79) **ke 'glavi**

casa agradável

casa bonita

Para Ferraz (1979: 80) a ordem substantivo + adjetivo é uma influência morfossintática africana, no entanto, poderia também ser portuguesa, pois esta é também a ordem das palavras em português.

Outras características do são-tomense não se assemelham nem ao português nem às línguas africanas de substrato, são características gramaticais inéditas. Algumas dessas características podem ser observadas no comportamento dos verbos e dos numerais.

## 9.1. Verbos

Com os dois quadros de conjugação a seguir, podemos, a princípio, perceber algumas particularidades da conjugação verbal em sãotomense.

**Tabela 25:** conjugação verbal no presente do indicativo.

<b>Pessoa</b>	<b>Sãotomense</b>	<b>Português</b>
1 SG	'nga-tlaba	(eu) trabalho
2 SG	bo ka 'tlaba	(tu) trabalhas
3 SG	e ka 'tlaba	(ele) trabalha
1 PL	nõ ka 'tlaba	(nós) trabalhamos
2 PL	i'nãse ka 'tlaba	(vocês) trabalham
3 PL	i'ně ka 'tlaba	(eles) trabalham

**Tabela 26:** conjugação verbal no pretérito perfeito do indicativo.

<b>Pessoa</b>	<b>Sãotomense</b>	<b>Português</b>
1 sg	'n-tlaba	(eu) trabalhei
2 sg	bo 'tlaba	(tu) trabalhaste
3 sg	e 'tlaba	(ele) trabalhou
1 pl	nõ 'tlaba	(nós) trabalhamos
2 pl	i'nãse 'tlaba	(vocês) trabalharam
3 pl	i'ně 'tlaba	(eles) trabalharam

Em sãotomense, as variações em pessoa, número, e tempo (aspecto) não são marcadas com morfemas, como ocorre em português e no crioulo caboverdiano. As pessoas e o número são indicados com a presença do pronome anteposto ao verbo, o aspecto é apresentado por meio de um sistema de partículas. Nos exemplos a seguir, pode-se observar que a partícula 'ka' marca o aspecto (aoristo) e os advérbios 'todo dia' e 'amanhã' funcionam como marcadores de aspecto habitual e tempo futuro, para evitar a ambiguidade do aoristo.

(80) Zõ **ka 'landa** 'tudo dza.

João KA nadar todo dia.

João nada todo dia.

(81) Zõ **ka 'landa** ama'ɲa.

João KA nadar amanhã

João vai nadar amanhã.

Antes de observar o estudo do verbo feito por Ferraz, é importante definir e diferenciar tempos e aspectos. Para Valkhoff (1966: 106), não existem tempos em são-tomense, apenas aspectos. O autor vai mais além e vê essa característica como universal nos crioulos.

‘Aspectos’ podem ser definidos como “uma categoria usada na descrição gramatical dos verbos (juntamente com tempo e modo), referindo-se à maneira como a gramática marca a duração ou o tipo de atividade temporal denotada pelo verbo” (Crystal, 2000: 32).

Grosso modo, podemos dizer que em português, os verbos podem ser analisados em tempo, modo e aspecto, sendo, por exemplo, uma forma como “trabalhávamos” do pretérito imperfeito do indicativo, como:

Tempo: pretérito.

Aspecto: imperfeito.

Modo: indicativo.

Tradicionalmente, os elementos são descritos na ordem TMA (tempo, modo, aspecto). No português, um contraste aspectual que pode ser levantado, é o do “perfeito” versus “imperfeito” nos pretéritos. “No contraste perfeito/imperfeito, o primeiro geralmente se refere a uma ação conclusa, e o segundo exprime duração sem especificar a conclusão da ação. Para o aspecto, apenas a duração ou o tipo de atividade temporal importam. Note que nos exemplos do caboverdiano:

(82) **bu bendi**, *vendeste* (pretérito perfeito do indicativo).

**bu konxi**, *conheces* (presente do indicativo). Quint (2007-a).

Embora os dois verbos se traduzam por tempos diferentes em português, estão ambos no mesmo aspecto, o perfectivo ou perfeito, que expressa o fato de a ação ter sido concluída. No primeiro caso, se o receptor vendeu alguma coisa no momento exato em que o emissor está falando, o receptor já concluiu a ação de vender, efetivamente vendeu a coisa em questão. No segundo caso, se uma pessoa conheceu alguém, conheceu *completamente*, isso implica que *ainda o conhece* no momento da realização do discurso. O aspecto continua sendo perfectivo, embora, em português, o tempo seja outro. O par de exemplos do caboverdiano não pode ser repetido igualmente em sãotomense, devido às idiossincrasias desta língua, contudo, outros pares podem ser formados, como os do aspecto aoristo, traduzidos por três tempos diferentes em português. Em sãotomense, as frases do caboverdiano são traduzidas:

(83)	Caboverdiano	Sãotomense	Português
	<b>bu bendi</b>	<b>bo vẽ'de</b>	vendeste
	<b>bu konxi</b>	<b>bo ka kõe'se</b>	conheces

Em sãotomense há seis aspectos verbais:

a) o perfectivo, ou aspecto perfeito, de ações perfeitamente concluídas. Equivale ao pretérito perfeito expresso do português e é o mais simples de conjugar. Basta unir a forma infinitiva do verbo ao sujeito:

#### **Sujeito<sup>49</sup> + verbo principal**

(84) **zõ 'landa.**

João nadar.

João nadou.

(85) **e 'gɔɓ 'gatu**

3SG procurar gato

Ele procurou pelo gato.

---

<sup>49</sup> “Noun phrase” no original.

## b) Aoristo (Presente habitual, futuro, condicional)

Do grego *aóristos*, significa “indefinido”. É o nome dado à “forma que o verbo toma para indicar que uma ação passada é vista independentemente de noções aspectuais, como, p. ex., ter sido completada ou não, estar repetida ou não, ser duradoura ou não” (Dicionário Aurélio). O aoristo, essa forma indefinida do verbo, é formado com a partícula aspectual KA, de origem controversa. Valkoff (1966: 105) menciona a existência de “ka” em outros crioulos de base europeia, não necessariamente portuguesa, como o crioulo francês da Antilhas, com o mesmo valor que em são-tomense. No crioulo de base francesa do Haiti e no crioulo de base holandesa das Ilhas Virgens, “ka” possui um valor diferente do valor em são-tomense. Para Valkoff, a presença do prefixo verbal “ka” em quase todos os crioulos é uma herança da língua franca portuguesa que lhes serviu de base (cf. seção 2.8.1. A teoria monogenética). Para Ferraz (1979: 114), a partícula aspectual “ka” é uma herança do bini, já que nesta língua existe a partícula *ya*, usada no futuro e no condicional, como em são-tomense. Quint (2007a: 4), ao contrário, afirma que a partícula aspectual “ka” não é influência do substrato, já que também existe nos crioulos de base lexical francesa das antilhas, que têm substrato wolof e mandinga, enquanto os crioulos do Golfo da Guiné possuem substrato banto e edoide. Para Quint (2007a: 3) as partículas aspectuais “na”, do crioulo da Guiné Bissau; “dja”, dos crioulos da Ásia; “al” do caboverdiano; e “lo” do papiamentu e de crioulos da Ásia são originárias do português: “em + a”, “já”, “há-de” e “logo”, respectivamente. Quint não arrisca uma hipótese para a origem para a partícula aspectual “ka” do são-tomense, que encontra paralelos nos outros crioulos do Golfo da Guiné.

Não obstante a ambiguidade do aoristo, o aspecto, o tempo e o modo são percebidos pelo contexto (devido também à presença dos advérbios de tempo ‘todo dia’ e ‘amanhã’), e não por morfemas ou partículas, como se pode observar nos exemplos a seguir, alguns dos quais já mencionados, mas agora glosados:

### **Sujeito + KA + Verbo principal**

Presente habitual:

(86) **Zõ ka 'landa 'tudo dza.**

João KA nadar todo dia.

João nada todo dia.

Futuro (do presente):

(87) **ama'ŋa e ka 'kɔpla 'tudu kwa.**

Amanhã 3SG KA comprar tudo coisa

Amanhã ele comprará todas as coisas.

Condicional (Futuro do pretérito):

(88) **Nga bε fi nga te 'tempu**

1SG-KA ir se 1SG-KA ter tempo.

Eu iria se tivesse tempo.<sup>50</sup>

c) o presente progressivo, ou gerúndio presente. Indica algo realizado no momento da fala. Equivale ao gerúndio em português.

### **Sujeito + sa ka + verbo principal**

(89) **Zõ sa ka 'landa**

João SA KA nadar

João está nadando.

A forma SA KA possui as variantes fonológicas **ska** e **ŋka**. A origem da partícula, segundo Ferraz é de SA (infinitivo do verbo SER, do português) e a forma KA.

d) GERÚNDIO PASSADO: O aspecto progressivo no pretérito.

### **Sujeito + (j)tava + KA + verbo principal**

Serve para uma ação em progresso no passado:

---

<sup>50</sup> Ferraz se esqueceu de mencionar o Pretérito Imperfeito do Subjuntivo, que pelo exemplo, se forma pelo mesmo aspecto do aoristo.

- (90) **e 'tava ka ngungu'nu ku zõ**  
 3SG estava KA resmungar com João  
 Ele estava resmungando com João.

Ou uma prática habitual.

- (91) **e 'tava ka ta'sõ na'i**  
 3SG estava KA sentar aí.  
 Ele costumava se sentar aí.

A origem da partícula aspectual (**j**)**tava** é evidentemente portuguesa.

Há ainda a partícula aspectual “kia” com o significado de “estar prestes a”. Todos os exemplos de Ferraz são do tempo passado. Eis um deles:

- (92) **Ola e kia tasom ni jadlim, nola se mem suba kye.**  
 /ɔ'la e 'kia ta'sõ ni za'dlĩ, 'n-ɔla sɛ mẽ 'suba kye/  
 Quando 3SG estar-prestes-a sentar em jardim, em-hora que muito chuva cair.  
 Na hora em que ele ia sentar no jardim, começou a chover.

**Tabela 27:** quadro resumo da conjugação verbal em sãotomense.

Tempo /aspecto	Elementos entre sujeito e verbo		Exemplo	Tradução
Aoristo		KA	Zõ ka 'landa	João nada (todo dia). João vai nadar (amanhã). João nadaria (se tivesse tempo).
Perfeito			Zõ 'landa	João nadou.
Gerúndio Presente	SA	KA	Zõ sa ka 'landa.	João está nadando.
Gerúndio Passado	TAVA	KA	Zõ tava ka 'landa	João estava nadando. João costumava nadar.
Mais que perfeito	TAVA		Zõ tava 'landa	João tinha nadado.
imperativo			'landa	Nade!

(montado a partir das informações de Ferraz, 1979).

## 9.2. Numerais

A entrada dos numerais no léxico sãotomense se deu de duas formas: de 1 a 10, os numerais são originários das formas portuguesas correspondentes, tendo sofrido apenas as alterações fonológicas. A partir do numeral 11, o sãotomense utiliza o léxico



português colado sobre a estrutura de formação de numerais do kikongo, a principal língua de substrato, mas não em perfeita sintonia com sua morfologia africana, como será demonstrado a seguir.

**Tabela 28:** numerais.

	Português	Sãotomense	kikongo
1	Um, uma	' <b>ũa</b>	-mofi
2	Dois, duas	' <b>dosu</b>	-ole
3	Três	' <b>tlefi</b>	-tatu
4	Quatro	' <b>kwatlu</b>	-ya
5	Cinco	' <b>jinku</b>	-tanu
6	Seis	' <b>sefi</b>	-sambanu
7	Sete	' <b>sete</b>	nsambwadi
8	Oito	' <b>oto</b>	nana
9	Nove	' <b>nove</b>	vwa
10	Dez	' <b>defi</b>	kumi
11	Onze	<b>def-k-ũa</b> <sup>51</sup> <b>dez mais um</b>	kumi ye mofi dez mais um
12	Doze	<b>def-ku'dosu</b> <b>dez mais dois</b>	kumi ye z-ole dez mais dois
20	Vinte	' <b>dosu 'defi</b> <b>dois dez</b>	makumole < ma-kumi m-ole dez dois, i.e. dois dez
24	Vinte e quatro	' <b>dosu 'defi ku 'kwatlu</b> <b>dois dez mais quatro</b>	makumole ye ya dois dez mais quatro
30	Trinta	' <b>tlefi 'defi</b> <b>três dez</b>	makumatatu < ma-kumi ma-tatu dez três, i.e. três dez
40	quarenta	' <b>kwatlu 'defi</b> <sup>52</sup> <b>quatro dez</b>	makumaya < ma-kumi ma-ya dez quatro, e.e. quatro dez

Tabela montada e reorganizada com base nas informações traduzidas de Ferraz (1979: 72-73).

Em kikongo, os numerais de 1 a 6 possuem diferentes e diversos prefixos de concordância, por se tratar de língua banta, mas os de 7 a 10 são invariáveis. A partir do 11, a formação é feita com base nos numerais de 1 a 10.

<sup>51</sup> Pode-se ainda deduzir que as formas do 11 ao 19 são etimologicamente portuguesas, já que 11 é, rigorosamente, “dez-com-um” e não “dez mais um”. Note-se ainda a queda do fonema [i] de *defi* para *def* em *def-k-ũa*.

<sup>52</sup> Devido à influência que o português vem exercendo no sãotomense até hoje, atualmente já se diz *kwalenta* para quarenta. Informação recebida oralmente do Prof. Dr. Gabriel Antunes de Araújo, que constatou *in loco* a nova palavra.

Pelo visto, a influência do português manteve-se mais uma vez forte no léxico, visto que se manteve para a formação de todos os numerais, enquanto a língua de substrato auxiliou na estrutura morfológica. Apesar disso, o sistema de numerais do sãoatomense não pode ser considerado um somatório da morfologia do kikongo com o léxico do português, visto que, em kikongo, a forma para o numeral 20 é “makumole” de “ma-kumi m-ole”, estrutura diferente da sãoatomense. Embora Ferraz não indique o significado de “ma”, trata-se de um prefixo de classe, apesar de afirmado anteriormente ser o numeral dez invariável em kikongo. Desta forma, uma glosa literal pode ser feita:

(93) Ma-kumi m-ole  
      PREF-dez PREF-dois  
      Vinte

O que não faz o falante confundir com doze, pois este é:

(94) Kumi ye z-ole  
      Dez mais PREF-dois

Ferraz (1979:73) mesmo indica “tens (*sic*) two, i.e. two tens”. Apesar disso, a palavra vinte, em sãoatomense, não é  $d\epsilon f$ -dosu, inconfundível com  $d\epsilon f$ -ku  ${}^1$ dosu (doze), como seria lógico se a cópia da morfologia kikongo fosse perfeita, mas sim  ${}^1$ dosu  ${}^1$  $d\epsilon f$ i, diferente do kikongo dez-dois e do português.

O sistema de numerais sãoatomense não é, portanto, nem português, nem kikongo, muito menos um somatório de ambos. Trata-se de um elemento inédito na língua.

## 10. CONSIDERAÇÕES FINAIS

Na fonologia, as influências das línguas africanas quicongo e bini no sãotomense podem ser observadas em fenômenos como o lambdacismo, por não haver [r] em quicongo (seções 3.1 e 6.2.4); a monotongação de [ej] em [e], por não existirem ditongos em quicongo; na palatalização e despalatalização das consoantes dentais e sibilantes, por haver fenômeno idêntico em quicongo (seções 3.1 e 6.2.3).

As influências da língua portuguesa na fonologia se observam pela confusão entre [v] e [b] em palavras em que havia tal variação no século XVI (seção 4.2.1.4); pela presença de fonemas como [ɾ], [ɲ] e [ʎ], inexistentes nas línguas africanas de substrato (seção 4.2.1.1); pela ausência de tons em sãotomense, embora existam em principense e angular; pela presença de [ʃ] e [ʒ] no quadro de fonemas e não de fones (seção 6.3); pelo rotacismo no processo atual de descrioulização (seção 7.3.1) e pela palatalização de [s] e [z] em coda de sílabas (seção 4.2.1.3) sendo que esta última é melhor considerada uma influência posterior à formação do crioulo.

É difícil afirmar, contudo, se é influência africana ou portuguesa a simplificação do sistema silábico para CV em grande maioria das sílabas, já que, embora as palavras sãotomenses sejam majoritariamente de estrutura CV, tal sistema silábico também é bastante comum em português. A presença das vogais nasais em sãotomense, já que, embora sejam inexistentes em quicongo, existiam em bini; a monotongação de [ow] para [o] e até mesmo a monotongação de [ej] para [e], já que, embora também existam em português, não é fenômeno geral, tal como ocorrem em sãotomense (seção 4.2.1.9). Sobre essas características existentes tanto na língua de superstrato quanto nas línguas de substrato, é bem provável que tal se deva a uma contribuição simultânea do substrato e do superstrato, e não de apenas uma isoladamente.

Na morfologia, as influências apresentadas também são tanto de substrato quanto de superstrato. O sistema de afixação do sãotomense não segue o sistema de classes banto, tampouco o sistema de plural português herdado do latim, mas sim o sistema de plural bini. O particípio dos verbos, contudo, é feito à maneira portuguesa, com acréscimo do sufixo “do”.

A formação do feminino, contudo, parece ser autêntica, não sendo cópia nem do sistema africano nem português. Também são autênticos o sistema verbal e dos numerais.

O léxico é formado em sua maioria por palavras portuguesas, o que é óbvio por se tratar de um crioulo de superstrato português, mas com um grande número de vocábulos africanos, até mesmo na lista-Swadesh.

As contribuições da língua portuguesa não são visíveis apenas no léxico, mas também na fonologia e na morfologia, assim como as influências africanas se mostram nas outras áreas da língua. Todas as línguas de substrato como a língua de superstrato contribuíram para a constituição do são-tomense, na fonologia, morfologia e léxico, embora em cada área tenha havido mais contribuições de uma língua do que de outra.

## 11. REFERÊNCIAS BIBLIOGRÁFICAS

- Agostinho, Ana L. S. & Araújo, Gabriel A. 2007. *Um estudo estatístico sobre as oxítonas no português*. Signótica 19.
- Angenot, J. P. 2007. A família linguística bantu: sua classificação e seus classificadores. Guajará-mirim: Universidade Federal de Rondônia. Mimeo, 18 p.
- Araújo, Gabriel Antunes. 2007. Empréstimos recentes do português, variação fonética e a sílaba na língua são-tomense da Ilha de São Tomé. In: *Papia*. Brasília.
- Arends et alii. 1995. *Pidgins and creoles an introduction*. Amsterdam: John Benjamins Publishing Company.
- Asopgelibre, 2007. *Obiang convierte al portugués en tercer idioma oficial para entrar en la comunidad lusófona de naciones*. [online], [acessado em 2 de nov. de 2007]. Disponível em: <<http://www.guinea-ecuatorial.net/ms/main.asp?cd=ni5393>>.
- Bagno, Marcos. 2006. *A norma oculta: língua & poder na sociedade brasileira*. São Paulo: Parábola Editorial. 5a edição.
- Barrena, Rvdo. P. Natalio. 1957. *Gramatica Annobonesa*. Madrid: Consejo Superior de Investigaciones Cientificas.
- Bennett, Patrick R. & Sterk, Jan P. (1977). *South Central Niger-Congo: A reclassification*. Studies in African Linguistics, 8, 241–273.
- Beyer, Klaus. 1986. *The Aramaic language: its distribution and subdivisions*. Göttingen: Vandenhoeck und Ruprecht.
- Bickerton, D. 1975. *Dynamics of a Creole System*. Cambridge Cambridge University Press.
- Bloom, L. 1970. *Language Development: Form and Function in Emerging Grammars*. Cambridge: MIT Press.
- Bollée, A. 1977. *Pidgins und kreolish Sprachen*. Studium Linguistik. Apud: Couto, Hildo H. 1994. *Introdução ao estudo das línguas crioulas e pidgins*. Brasília: Editora da UnB.
- Bonvini, E. & Petter, M. M. T. 1998. *Portugais du Brésil et langues africaines in Langages L'hyperlangue brésilienne*. Paris: Larousse.
- Bowerman, M. 1973. *Early Syntactic Development: A Cross-linguistic Study, with special reference to Finnish*. Cambridge: CUP.
- Boxer, Charles R. 2002. *O império marítimo português: 1415-1825*. 1ª reimpressão. São Paulo: Companhia das Letras.

- Buescu, Maria L. C. 1975. *A Gramática de Linguagem Portuguesa de Fernão de Oliveira. Introdução, leitura actualizada e notas por M. L. C. Buescu*. Lisboa: Imprensa Nacional – Casa da Moeda.
- Camara Jr., J. M. 1968. *Dicionário de filologia e gramática*. Rio de Janeiro: J. Ozon Editor.
- \_\_\_\_\_. 1975. *História e estrutura da língua portuguesa*. Rio de Janeiro: Padrão.
- \_\_\_\_\_. 1998. *Estrutura da Língua Portuguesa*. Petrópolis: Vozes.
- Camões, Luiz V. 1572. *Os Lusíadas*. Edição fac similada, publicada em 1972 pelo MEC.
- Caminha, Pero Vaz. 2002. *Carta de Pero Vaz de Caminha*. São Paulo: Martin Claret.
- Castro, Ivo. 1991. *Curso de história da língua portuguesa*. Lisboa: Universidade Aberta.
- Castro, Yeda P. 2006. *A influência das línguas africanas no português brasileiro*. [online], [acessado em 20 de novembro de 2006]. Disponível em: <<http://www.smecc.salvador.ba.gov.br/documentos/linguas-africanas.pdf>>
- Cia Factbook. *Sao Tome and Principe*. [online], [acessado em 20 de janeiro de 2008]. Disponível em: <<https://www.cia.gov/library/publications/the-world-factbook/geos/tp.html>>.
- Cintra, Luis F. L. 1963. Les anciens textes non littéraires, classement et bibliographie. *Revue de Linguistique Romane*, 27, pp. 40-58.
- Clements, G. N. & Keyser, S. J. 1983. *CV Phonology: A Generative theory of the Syllable*. Cambridge: Mit Press.
- Coene, A. 1960. *Kikongo, Notions grammaticales, Vocabulaire Français – Kikongo – Néerlandais – Latin*. Imprimerie Mission Catholique Tumba. Edição fac similada. Disponível em: <http://www.nekongo.org/akongo/>
- Corder, S. P. 1978. The Language Learner's Language. In: Richards, I. (ed.) *Understanding Second and Foreign Language Learning. Issues and Approaches*. Nova York: Newbury House.
- Coseriu, Eugenio. 2000. Introdução. In: Torres, Amadeu & Assunção, Carlos. 2000. *Gramática da Linguagem Portuguesa. Edição Crítica, Semidiplomática e Anastática*. Braga: Barbosa & Xavier, Ltda.
- Coutinho, Ismael de Lima. 1976. *Pontos de Gramática Histórica*. Rio de Janeiro: Ao Livro Técnico. 7ª edição.

- Couto, Hildo H. 1994. *Introdução ao estudo das línguas crioulas e pidgins*. Brasília: Editora da UnB.
- \_\_\_\_\_. 1999. *Contato Interlinguístico: Da Interação à Gramática*. [online]. [Acessado em: 19 de janeiro de 2008]. Disponível em: <<http://www.unb.br/il/let/crioul/contato.htm>>.
- Creissels, Denis. 1994. *Aperçu sur les structures phonologiques des langues négro-africaines*. Grenoble: Ellug, Deuxième édition.
- Crofts, Marjorie. 2006. *Complexidades sociolinguísticas ocorrentes na tradução da literatura de autoria indígena para a língua nacional*. [online]. [Acessado em: 4 de março de 2007]. Disponível em: <<http://www.sil.org/americas/brasil/PUBLICNS/LING/CompTrad.pdf>>.
- Crystal, David. 2000. *Dicionário de Linguística e Fonética*. Rio de Janeiro: Jorge Zahar Editor.
- Cuesta, Pilar Vasquez, & M.A.M Luz. 1961 *Gramática portuguesa*. Madrid: Gredos.
- Cunha, Celso. 1980. O protocrioulo português e sua universalidade nos séculos XVI, XVII e XVIII. In: Cunha, Celso. *Língua, Nação e Alienação*. Rio de Janeiro: Nova Fronteira.
- Dereu, Léon. 1955. *Cours de kikongo*. Bruxelles: Maison d'Éditions ad. Wesmael Charlier Namur.
- Dumestre, Gérard. 1998. Les idéophones: le cas du bambara. *Faits de Langues, Revue de Linguistique*, n-º. 11-12, Les Langues d'Afrique subsaharienne. Paris: Ophrys.
- Edo Language Dictionary Online. 2008. [online], [acessado em 30 de janeiro de 2008]. Disponível em: <<http://www.edofolks.com/words/index.htm>>.
- Eweka, Iyi. 2008. *Edo from Sao Tome and Principe*. [online]. [Acessado em: 27 de janeiro de 2008]. Disponível em: <<http://www.edo-nation.net/saotome.htm>>.
- Ferreira, Aurélio B. H., 2006. *Novo dicionário Aurélio da Língua Portuguesa*, versão 5.0 para computador, correspondente à 3ª edição, 1ª impressão da Editora Positivo, revista e atualizada do Aurélio Século XXI.
- Ferraz, Luiz I. 1976. A origem e desenvolvimento de quatro crioulos portugueses do Golfo da Guiné. In: *Revista Brasileira de Linguística*, 3 (2): páginas 22-78.
- \_\_\_\_\_. 1979. *The creole of São Tomé*. Johannesburg: Witwatersrand University Press.
- Galves, Charlotte. 2006. *A língua das caravelas: Periodização do português europeu e origem do português brasileiro*. Unicamp. [online]. [Acessado em: 2 de dezembro

de 2007]. Disponível em: <[http://www.ime.usp.br/~tycho/prfpml/relat\\_2006/origem.pdf](http://www.ime.usp.br/~tycho/prfpml/relat_2006/origem.pdf)>.

Gonçalves Dias, 1969. *Antologia Poética*. 5ª edição. Rio de Janeiro: Agir.

Gordon, Raymond G., editor. 2005a. *Ethnologue: languages of the world*. 15th ed. Dallas: SIL International. *Edo: a language of Nigeria*. [online], [acessado em 15 de abril de 2007]. Disponível em: <[http://www.ethnologue.com/show\\_language.asp?code=bin](http://www.ethnologue.com/show_language.asp?code=bin)>.

\_\_\_\_\_. 2005b. *Languages of São Tomé e Príncipe* [online], [acessado em 15 de abril de 2007]. Disponível em: <[http://www.ethnologue.com/show\\_country.asp?name=ST](http://www.ethnologue.com/show_country.asp?name=ST)>.

\_\_\_\_\_. 2005c. *Kongo: a language of Democratic Republic of Kongo*. [online], [acessado em 15 de agosto de 2007]. Disponível em: <[http://www.ethnologue.com/14/show\\_language.asp?code=KON](http://www.ethnologue.com/14/show_language.asp?code=KON)>.

\_\_\_\_\_. 2005d. *Language family trees. Creole, Portuguese based*. [online], [acessado em 15 de agosto de 2007]. Disponível em: <[http://www.ethnologue.com/14/show\\_family.asp?subid=942](http://www.ethnologue.com/14/show_family.asp?subid=942)>.

\_\_\_\_\_. 2005e. *Kituba: a language of Democratic Republic of Kongo*. [online], [acessado em 15 de agosto de 2007]. Disponível em: <[http://www.ethnologue.com/14/show\\_language.asp?code=KTU](http://www.ethnologue.com/14/show_language.asp?code=KTU)>.

\_\_\_\_\_. 2005f. *Language family index* [online], [acessado em 1 de janeiro de 2008]. Disponível em: <[http://www.ethnologue.com/14/family\\_index.asp](http://www.ethnologue.com/14/family_index.asp)>.

\_\_\_\_\_. 2005g. *Ethnologue language family index* [online], [acessado em 1 de janeiro de 2008]. Disponível em: <[http://www.ethnologue.com/family\\_index.asp](http://www.ethnologue.com/family_index.asp)>.

Graham, Steve & Graham, Trina. 2004. *West Africa lusolexed creoles word list file documentation* [online], [acessado em 15 de abril de 2007]. Disponível em: <<http://www.sil.org/silesr/abstract.asp?ref=2004-012>> e <<http://www.sil.org/silesr/2004/images/Wordlists.pdf>>.

Günther, Wilfred. 1973. *Das portugiesische Kreolisch der Ilha do Príncipe*. Berlim: Marburg an der Lahn.

Hall, Robert A. 1966. *Pidgin and creole languages*. Ithaca, N.Y.: Cornell University Press.



- Hartell, Rhonda L. 1990. *Alphabets of Africa*. Dakar: Unesco. Edição fac simulada. [online]. [Acessado em: 27 de janeiro de 2008]. Disponível em: <[http://www.rosettaproject.org/archive/bin/ortho-1?page\\_view=image\\_view](http://www.rosettaproject.org/archive/bin/ortho-1?page_view=image_view)>.
- Heine, Bernd. 2006. *An areal perspective on Africa*. Addis Ababa: Mimeo.
- Henriques, Isabel C. 2000. *São Tomé e Príncipe: A Invenção de Uma Sociedade*. Lisboa: Vega.
- Houaiss, Antonio & Villar, Mauro. 2001. *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro: Objetiva.
- Igbineweka, Alex G. 2008-a. *Teach yourself edo language*. [online]. [Acessado em: 27 de janeiro de 2008]. Disponível em: <<http://www.edo-nation.net/edolang1.htm>>.
- \_\_\_\_\_. 2008-b. *The guosa language*. [online]. [Acessado em: 27 de janeiro de 2008]. Disponível em: <<http://www.dawodu.net/guosa1.htm>>.
- Katamba, F. 1993. *An introduction to phonology*. Londres: Longman, 6 impressão.
- Koseki, Kazuaki. 2007. *Palavras japonesas de origem portuguesa*. [online]. [Acessado em: 14 de agosto de 2007]. Disponível em: <<http://www.fcsh.unl.pt/ilnova/japones.htm>>.
- Ladefoged, Peter. 1993. *Vowels and consonants*. 2ª edição. Blackwell Publishers. [material multimídia: CD-Room].
- Lopes, L. P. M. 2001. *Oficina de linguística aplicada: a natureza social e educacional dos processos de ensino/aprendizagem de línguas*. Campinas: Mercado de Letras. 3ª reimpressão.
- Lucchesi, D. & Baxter, A. 2006. *Projeto Vertentes do Português Rural do Estado da Bahia*. [online]. [Acessado em: 20 de julho de 2007]. Disponível em: <[www.vertentes.ufba.br/helvecia.htm](http://www.vertentes.ufba.br/helvecia.htm)>.
- Lumwamu, François. 1973. *Essai de morphosyntaxe systématique des parlers kongo*. Paris: Peeters Publishers. Edição fac simulada. Disponível em: <<http://books.google.com>>.
- Macek, J. F. 2007. *Kiswahili, uma língua altamente adaptável – Empréstimos do português no kiswahili*. Hand-out de comunicação no 3º Mini Enapol Gela, realizado na Casa de Cultura Japonesa em 13 de abril de 2007. São Paulo: Mimeo.
- Mané, Djiby. 2003. Estudo comparativo dos crioulos portugueses do Golfo da Guiné: o são-tomense, o angolano, principense e anobonês. *Papia* 13, p. 31-36.
- \_\_\_\_\_. 2007. Os crioulos portugueses do Golfo da Guiné: quatro línguas diferentes ou dialetos de uma mesma língua? Tese de Doutorado, Brasília: UnB.

- Marroquim, M. 1934. *A Língua do Nordeste (Alagôas e Pernambuco)*. São Paulo: Companhia Editora Nacional. Apud: Ferraz, Luiz I. 1979. *The creole of São Tomé*. Johannesburg: Witwatersrand University Press.
- Martins, Vicente. 2006. *Fonologia: as consoantes do português*. [online]. [Acessado em 24 de janeiro de 2008]. Disponível em: <<http://inforum.insite.com.br/8920/2898593.html>>.
- Maurer, Philippe. 1995. *L'angolar: Um créole afro-portugais parlé à São Tomé*. Hamburgo: Helmut Buske Verlag.
- Mendonça, Renato. 1973. *A Influência Africana no Português do Brasil*. Rio de Janeiro: Civilização Brasileira.
- Mühlhäusler, Peter. 1986. *Pidgin and creole linguistics*. Oxford: Blackwell.
- Muysken, Pieter & Smith, Norval. 1995. The study of pidgin and creole languages. In: Arends, Jacques, Pieter Muysken & Norval Smith. 1995. *Pidgins and creoles: an introduction*. Amsterdam: John Benjamins.
- Navarro, E. A. 1999. *Método moderno de tupi antigo*. Petrópolis: Vozes. 2ª edição.
- Negreiros, A.. 1895. História Ethnographica da Ilha de S. Thomé. Lisboa. Apud: Ferraz, L. 1979. *The creole of São Tomé*. Johannesburg: Witwatersrand University Press.
- Nicola, J. & Infante, U. 1994. *Gramática Contemporânea da Língua Portuguesa*. São Paulo: Scipione. 13ª edição.
- Okoudowa, Bruno. 2005. Descrição preliminar de aspectos da fonologia e da morfologia do lembaama. Dissertação de Mestrado, São Paulo: USP.
- Oliveira, Fernão de. 1536. (1988). *Grammatica da lingoagem portuguesa* (Gramática da linguagem portuguesa). Edição fac-similada. Lisboa: Biblioteca Nacional.
- Oliveira, Marcia S. D. 2007. *Mini-curso: Línguas Africanas: seu conhecimento e pesquisa*. [online]. [Acessado em: 29 de dezembro de 2007]. Disponível em: <[http://www.fflch.usp.br/dlcv/m\\_oliveira/textos/pdf/Mini-curso\\_UFMS.pdf](http://www.fflch.usp.br/dlcv/m_oliveira/textos/pdf/Mini-curso_UFMS.pdf)>.
- Otero, Edgardo. 2006. *A Origem dos Nomes dos Países*. São Paulo: Panda Books, 1ª edição.
- Payne, Thomas E. 1997. *Describing Morphosyntax*. [online]. [Acessado em: 26 de janeiro de 2008]. Disponível no google books: <<http://books.google.com.br>>.
- Pereira, R. P. 2007. *A construção do nacionalismo em São Tomé e Príncipe*. [online]. [Acessado em: 29 de dezembro de 2007]. Disponível em: <<http://www.klepsidra.net/klepsidra13/saotome.htm>>.
- Petter, M. M. T. 2005. *Línguas africanas no Brasil*. In: Gragoatá. Niterói: UFF.

- \_\_\_\_\_. 2006. *Sistemas vocálicos*. Handout de aula na disciplina Aspectos da tipologia das línguas negro-africanas, ministrada na Faculdade de Letras da Universidade de São Paulo no segundo semestre de 2006. São Paulo: USP.
- Pinker, Steven. 2002. *O instinto da linguagem: como a mente cria a linguagem*. São Paulo: Martins Fontes.
- Post, Marilke. 1995. Fa d'Ambu. In: Arends, Jacques; Muysken, Pieter; Smith, Norval. 1995. *Pidgins and creoles: an introduction*. Amsterdam: John Benjamins.
- Queiroz, Rita C. R. 2007. *Para que editar? A memória linguística preservada através dos textos*. [online]. [Acessado em: 19 de janeiro de 2008]. Disponível em: <<http://www.filologia.org.br/revista/38/08.html>>.
- Quint, Nicolas. 2007a. Os sistemas TAM nos crioulos de base lexical portuguesa. Trabalho apresentado no minicurso de mesmo nome, ministrado pelo Prof. Dr. Nicolas Quint, em visita à USP, em 10 de abril de 2007. São Paulo: mimeo.
- \_\_\_\_\_. 2007b. As principais abordagens relativas à questão da origem das línguas crioulas, com especial atenção aos crioulos afro-europeus. Trabalho apresentado no 3º Mini Enapol GELA, realizado na Casa de Cultura Japonesa em 13 de abril de 2007. São Paulo: mimeo.
- Quintão, J. L., 1934. *Gramática de Kimbundo*. 1ª ed. Edições Descobrimento.
- Raposo, Renato & Vaz, Francine. 2002. *Aquisição da linguagem*. [online], [acessado em 19 de janeiro de 2008]. Disponível em: <<http://www.nce.ufrj.br/GINAPE/publicacoes/trabalhos/RenatoMaterial/aquisicao.htm>>.
- Rauta, Aurelio. 1947. *Gramatica Rumana*. Madrid: Universidad de Salamanca.
- Ribeiro et alii. 2007. *Períodos da Língua Portuguesa*. [online]. [Acessado em: 2 de dezembro de 2007]. Disponível em: <[http://www.ime.usp.br/~tycho/participants/psousa/cursos/materiais/mesa\\_maceio\\_intro\\_geral.pdf](http://www.ime.usp.br/~tycho/participants/psousa/cursos/materiais/mesa_maceio_intro_geral.pdf)>.
- Rio de Janeiro. 2008. Encyclopædia Britannica. *Ultimate Reference Suite*. Chicago: Encyclopædia Britannica.
- Rodrigues, Aryon. 1986. *Línguas brasileiras*. São Paulo: Loyola. Citado em: Agostinho, Ana L. S. & Araújo, Gabriel A. 2007. *Um estudo estatístico sobre as oxítonas no português*. Signótica 19.
- Rougé, Jean-Louis. 2004. *Dictionnaire étymologique des créoles portugais d'Afrique*. Paris: Karthala.
- \_\_\_\_\_. 2006. *Notas do minicurso "Interações e crioulações"* ministrado pelo prof. Dr. Jean-Louis Rougé da Universidade de Orleans, na Faculdade de

Filosofia, Letras e Ciências Humanas da USP - Universidade de São Paulo, no período de 22 a 25 de agosto de 2006.

- Said Ali, Manuel. 1921. *Lexicologia do Português Histórico*. [Acessado em: 20 de janeiro de 2008]. Disponível em: <<http://ciberduvidas.sapo.pt/antologia.php?rid=582>>.
- Scherre, Maria M. P & Naro, Anthony J. 2001. Sobre as origens estruturais do português brasileiro: crioulização ou mudança natural? In: *Papia*, 11, p. 40.50.
- Silva, Rosa V. M. 1991. *O português arcaico: fonologia*. São Paulo: Contexto.
- Silva, Thais C. 1999. *Fonética e fonologia do português*. São Paulo: Contexto.
- Smith, Jean-François. 2007. *São Tomé-et-Príncipe*. [online]. [Acessado em: 20 de novembro de 2007]. Disponível em: <<http://www.tlfq.ulaval.ca/AXL/AFRIQUE/Sao-Tome-Principe.htm>>.
- Stillings, Neil A. 1987. *Cognitive Science: an introduction*. Cambridge: Massachusetts Institute of Technology.
- Teyssier, Paul. 2004. *História da Língua Portuguesa*. São Paulo: Martins Fontes.
- Thomas, N. W. 1910. *Anthropological Report on the Edo-Speaking people of Nigeria*. London: Harrison and sons. Edição fac similada. [online]. [Acessado em: 27 de janeiro de 2008]. Disponível em: <[http://www.rosettaproject.org/archive/bin/phon-3/?b\\_start:int=1&page\\_view=image\\_view](http://www.rosettaproject.org/archive/bin/phon-3/?b_start:int=1&page_view=image_view)>.
- Thompson, R. W. 1959. A note on some possible affinities between the creole dialects of the Old World and those of the New. In: *Proceedings of the conference on creole language studies*. London: MacMillan. p. 107. Apud: Cunha, Celso. 1980. O protocioulo português e sua universalidade nos séculos XVI, XVII e XVIII. In: Cunha, Celso. *Língua, Nação e Alienação*. Rio de Janeiro: Nova Fronteira.
- Tim, Wallace. 2008. *Phonetics and morphophonemics*. NC State University. Revised in January 2008. [online]. [Acessado em: 29 de janeiro de 2008]. Disponível em: <http://www4.ncsu.edu/~twallace/ANT254%20HW7.htm>
- Valkhoff, M. F. 1966. *Studies in portuguese and creole*. Johannesburg: Witwatersrand University Press.
- Valente, P. José F. 1964. *Gramática umbundu: a língua do centro de Angola*. Lisboa: Junta de investigações do ultramar.
- Vasconcelos, José L. 1959. *Lições de Filologia portuguesa*. Rio de Janeiro: Livros de Portugal.

- Whinnom, Keith. 1956. *Spanish contact vernacular in the Phillipine Islands*. Hong Kong: University Press. Apud: Cunha, Celso. 1980. O protocioulo português e sua universalidade nos séculos XVI, XVII e XVIII. In: Cunha, Celso. *Língua, Nação e Alienação*. Rio de Janeiro: Nova Fronteira.
- Wikipedia. 2006. *Kongo language*. [online]. [Acessado em: 20 de dezembro de 2006]. Disponível em: <[http://en.wikipedia.org/wiki/Kongo\\_language](http://en.wikipedia.org/wiki/Kongo_language)>.
- \_\_\_\_\_. 2007. *Kikongo/Kituba*. [online]. [Acessado em: 25 de novembro de 2007]. Disponível em: <<http://kg.wikipedia.org/wiki/Kituba>>.