



UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LÍNGUA E CULTURA

**VARIAÇÃO NA CONCORDÂNCIA VERBAL NO PORTUGUÊS DA
COMUNIDADE DE ALMOXARIFE, SÃO TOMÉ (ÁFRICA)**

THAMIRIS SANTANA COELHO

Salvador

2015

THAMIRIS SANTANA COELHO

**VARIAÇÃO NA CONCORDÂNCIA VERBAL NO PORTUGUÊS DA
COMUNIDADE DE ALMOXARIFE, SÃO TOMÉ (ÁFRICA)**

Dissertação apresentada ao Curso de Mestrado em Letras,
do Programa de Pós-Graduação em Língua e Cultura da
Universidade Federal da Bahia – UFBA, como requisito
parcial para obtenção do grau de Mestre em Letras.

Orientador: Prof. Dr. Alan N. Baxter

Salvador

2015

AGRADECIMENTOS

Apesar do percurso, por vezes, solitário que é a trajetória acadêmica, há algumas pessoas que o tornam humano. Por isso, acho justo tornar público meu sincero agradecimento às pessoas que foram e são de grande importância para que eu chegasse nessa etapa marcante.

Ao meu orientador, Alan N. Baxter, pela dedicação em transmitir o seu conhecimento acerca do estudo linguístico, pela constante paciência e humildade, ao mostrar que a inteligência deve estar atrelada à afabilidade.

Ao professor Dante Lucchesi por ser sempre solícito e por tornar fascinante o debate sociolinguístico. À professora Eliana Pitombo por ter me apresentado à pesquisa científica e por ter me incentivado a fazer o mestrado.

Aos meus amigos de jornada na Universidade Federal da Bahia: Adilson, Flávia, Isabel, Illa, Lidiane, Marília, Nélio e Rose, pelo companheirismo e encorajamento mútuo nos diversos momentos em as forças se esvaíam e, pelas alegrias vivenciadas nesse período.

À Joelaine e Tayana pelo acolhimento em Salvador. À Larissa por ser um anjo bom sempre que precisei.

À minha família por todo incentivo. Em especial, à minha mãe Adailza, pela motivação que me fortalecia todos os dias. Agradeço ao meu pai Renato, aos meus irmãos Sillas, Paulo Natã e Naara por todo apoio.

Ao meu marido, Daniel, pela cumplicidade, por entender meus momentos de quietude para estudar e, principalmente, pelo amor que as palavras não são capazes de descrever.

A Deus, por sua infinita graça e misericórdia, pois sem Ele nada disso seria possível.

RESUMO

Esta dissertação estuda a variação na concordância verbal junto à primeira pessoa do plural (P4) do português falado na comunidade Almojarife (PA), localizada em São Tomé, na África, com o objetivo de discutir o status sociolinguístico dessa variedade africana do português no âmbito da Sociolinguística Variacionista Laboviana. A discussão de estudos anteriores sobre a variável de P4 em variedades do Português Brasileiro (PB), fornece a matriz teórico-metodológica para o presente estudo, além dos pontos de comparação para os resultados da análise eventual. Parte-se do princípio de que as referidas variedades da língua portuguesa caracterizam-se por uma redução na sua morfologia flexional, motivada pelo contato entre línguas e os processos de aquisição de segunda língua. Acredita-se, portanto, que os fatores sociais, a história e os aspectos relacionados com a Gramática Universal fizeram emergir na comunidade de Almojarife uma variedade de português reestruturado. Na análise dos dados, o estudo recorre ao pacote estatístico GoldVarb (2001), que permite identificar as variáveis linguísticas e extralinguísticas que influenciam a concordância P4 nessa variedade de português. Foram analisadas 18 entrevistas, sendo 9 de informantes do sexo masculino e 9 do sexo feminino. Constatou-se que a regra de concordância verbal de primeira pessoa do plural é aplicada numa frequência de 73, 6% contra 26, 4% de não concordância nos dados encontrados, contrariando a hipótese geral de que os falantes fariam pouco uso das marcas de plural nos verbos. Em relação à gramática da comunidade, a análise quantitativa revelou que a concordância P4 é mais empregada quando o sujeito não é realizado e quando a diferença morfofonológica entre o singular e o plural do verbo é mais saliente. Desde a perspectiva do encaixamento social da variação, os dados revelaram que os mais jovens aplicaram mais a regra de concordância enquanto que os homens tendem a realizar mais a concordância do que as mulheres. A comparação dos resultados com os dos estudos brasileiros permitiu concluir que o PA, que tem como substrato o crioulo santome, manifesta padrões de variação semelhantes àsquelas de variedades do PB, que surgiram de uma situação de contato entre o português e línguas africanas, e de transmissão linguística irregular.

Palavras-chave: Concordância verbal; variação; primeira pessoa do plural; Português de São Tomé; Sociolinguística.

ABSTRACT

This dissertation studies variation in verbal agreement in the first-person plural (P4) of the Portuguese spoken in the Almoxarife community (PA), located in São Tomé, Africa. The purpose is to discuss the sociolinguistic status of this characteristic of this variety of African Portuguese within the framework of Labovian Variationist Sociolinguistics. A discussion of previous studies on the P4 variable in varieties of Brazilian Portuguese (BP) are discussed, provides the theoretical and methodological matrix for the present study, as well as points of comparison for the results of the eventual analysis. The study is based on the premise that these varieties of Portuguese are characterized by a reduction in inflectional morphology, caused by language contact and second language acquisition. As such, it is assumed that social and historical factors, and aspects of Universal Grammar resulted in the emergence of a variety of restructured Portuguese in the Almoxarife community. For the data analysis, the study uses the GoldVarb (2001) statistical package, which permits identification of the linguistic and extralinguistic variables that influence P4 agreement in this variety of Portuguese. 18 interviews were analyzed: 9 of male respondents and 9 of females. It was found that verb agreement rule of first person plural is applied at a frequency of 73, 6% in contrast with a non-application frequency of 26, 4%, contrary to the general hypothesis adopted in the study, that the speakers would make little use of plural agreement on the verb. In relation to the community grammar, the quantitative analysis revealed that P4 agreement is most favoured when the subject is absent and when the morphophonological difference between the singular and plural forms of verb is most salient. With regard to the social embedding of the variation, the data revealed that younger speakers most favoured the agreement rule while, in general, the men favoured the rule more than did the women. Comparison of the results with those of Brazilian studies revealed that PA, which has São Tomé Creole Portuguese as its substrate, displays patterns of variation similar to those found in varieties of BP, the latter being linguistic varieties which emerged from a situation of contact between Portuguese and African languages, and through irregular language transmission.

Key-words: subject-verb agreement; first-person plural variation; São Tomé Portuguese; sociolinguistics.

LISTA DE FIGURA

FIGURA 1. Índice demográfico de Santana e Almoxarife	45
--	----

LISTA DE GRÁFICOS

GRÁFICO 1. Distribuição da população por línguas faladas em São Tomé e Príncipe.....	21
GRÁFICO 2. Resultado geral do uso da CV-PA em Almoxarife.....	62
GRÁFICO 3. Inclinação da frequência de CV-P4 entre as faixas etárias na comunidade de Almoxarife em São Tomé e Príncipe.....	76

LISTA DE QUADROS

QUADRO 1. A expressão de pessoa e número com o verbo ‘saber’ em santome e português.....	22
QUADRO 2. Representação dos níveis de saliência fônica.....	25
QUADRO 3. Processos de variação no paradigma da flexão verbal no português brasileiro culto e no português popular do Brasil.....	30
QUADRO 4. Variáveis atuantes para o cancelamento da marca de número no verbo por variedade.....	33
QUADRO 5. Informantes de Almozarife: faixas etárias e seus aspectos sociais.....	50
QUADRO 6. Variáveis independentes processadas pelo VARBRUL.....	65

LISTA DE TABELAS

TABELA 1. Comparação dos níveis 1 e 4 de saliência fônica entre o PB e PST no cancelamento da CV-P6	33
TABELA 2. Comparação dos traços semânticos [+animado] e [-animado] do sujeito entre o PB e PST no cancelamento da CV-P6.....	34
TABELA 3. Comparação da faixa etária entre o PB e PST no cancelamento da CV-P6	34
TABELA 4. Frequência de aplicação da regra da CV-P4 na fala da comunidade de Almojarife em São Tomé.....	62
TABELA 5. Efeito da <i>realização e posição do sujeito</i> sobre a variável dependente.....	66
TABELA 6. Efeito da variável <i>saliência fônica</i> , sobre a variável dependente.....	69
TABELA 7. Frequência geral da variável <i>paralelismo formal</i>	71
TABELA 8. Efeito da variável <i>sexo</i> sobre a variável dependente.....	73
TABELA 9. Efeito da variável <i>faixa etária</i> sobre a variável dependente.....	75
TABELA 10. Efeito da variável <i>escolarização</i> sobre a variável dependente.....	78
TABELA 11. Cruzamento de <i>faixa etária</i> com <i>escolarização</i>	79
TABELA 12. Efeito da variável <i>estada fora da comunidade</i> sobre a variável dependente....	80

LISTA DE ABREVIATURAS

CV	concordância verbal
DAL	dispositivo de aquisição da linguagem
DLP	dados linguísticos primários
DOC	documentador
INF	informante
L1	primeira língua
L2	segunda língua
LA	língua alvo
P4	primeira pessoa do plural
P6	terceira pessoa do plural
PA	português de Almojarife
PB	português brasileiro
PE	português europeu
PEUL	Projeto de Estudos do Uso da Língua
PR	peso relativo
PST	português de São Tomé e Príncipe
R1	rodada 1
R2	rodada 2
R3	rodada 3
SV	sintagma verbal
SN	sintagma nominal
TMA	tempo, modo e aspecto

SUMÁRIO

INTRODUÇÃO	12
CAPÍTULO 1 – ENQUANDRAMENTO TEÓRICO-DESCRITIVO DA CONCORDÂNCIA VERBAL DE PRIMEIRA PESSOA DO PLURAL.....	16
1.1 O PERCURSO HISTÓRICO DO QUADRO LINGUÍSTICO DE SÃO TOMÉ E PRÍNCIPE.....	17
1.1.1 As línguas crioulas de São Tomé.....	18
1.1.2 O Português em São Tomé.....	20
1.2 PESQUISAS VARIACIONISTAS SOBRE A CONCORDÂNCIA VERBAL.....	24
1.3 AQUISIÇÃO E TRANSMISSÃO GERACIONAL.....	36
1.4 SÍNTESE.....	42
CAPÍTULO 2 – METODOLOGIA.....	43
2.1 A COMUNIDADE DE FALA EM FOCO.....	44
2.1.1 Características do português de Almojarife.....	46
2.1.2 Origem dos dados; estrutura do <i>corpus</i>.....	49
2.2 ENQUADRAMENTO TEÓRICO.....	52
2.2.1 Aspectos metodológicos.....	53
2.3 AS HIPÓTESES DO TRABALHO.....	56
2.4 SÍNTESE.....	60
CAPÍTULO 3 – ANÁLISE DOS DADOS.....	61
3.1 RESULTADOS GERAIS.....	62
3.2 ENCAIXAMENTO ESTRUTURAL.....	66
3.2.1 Variável <i>realização e posição do pronome sujeito</i>.....	66
3.2.2 Variável <i>saliência fônica</i>.....	69

3.2.3 Variável <i>paralelismo formal</i>	71
3.3 ENCAIXAMENTO SOCIAL	73
3.3.1 Variável <i>sexo do falante</i>	73
3.3.2 Variável <i>faixa etária do falante</i>	75
3.3.3 Variável <i>escolarização</i>	78
3.3.4 Variável <i>estada fora da comunidade</i>	80
3.4 SÍNTESE	81
CONSIDERAÇÕES FINAIS	82
REFERÊNCIAS	86

INTRODUÇÃO

Na busca por ampliar o campo de investigação acerca do português de São Tomé (PST), nesta dissertação apresentam-se os resultados empíricos de uma pesquisa sociolinguística sobre o fenômeno morfossintático da concordância verbal de primeira pessoa do plural (CV-P4) no português falado na comunidade de Almojarife (PA). Este trabalho tem a finalidade de cotejar os resultados obtidos com estudos feitos para o PB.

É irrefutável falar sobre CV sem citar o PB e o PE, pois a variação da CV se configurou como um fenômeno que evidencia a diferença morfossintática entre essas variedades do português (Naro e Scherre, 1993; 2007). Além disso, a CV também polariza a realidade sociolinguística do PB com uma norma culta e uma norma popular (BAXTER e LUCCHESI, 1997; LUCCHESI, 1998, 2000, 2001;).

Esses estudos originaram um debate acerca das motivações que determinam a variação da CV. Sumariamente, os debates teóricos giram em torno das seguintes hipóteses¹:

(i) a influência das línguas de substrato africano (GUY, 1981) que propõe que essa variação é decorrente de um ou mais substratos;

(ii) a deriva românica (NARO, 1981; SCHERRE, 1988) que defende a mudança linguística relacionada à fenômenos históricos internos da própria língua;

(iii) a transmissão linguística irregular (BAXTER, 1995; BAXTER e LUCCHESI, 1997) que sugere que a mudança na língua e a reestruturação se deve a um *input* de dados de português L2 (segunda língua) no processo de aquisição de L1 (primeira língua);

(iv) a formação das línguas reestruturadas assimilando as três proposições anteriores: a deriva, a transmissão linguística irregular e a influência das línguas crioulas (HOLM, 1988).

O comportamento da concordância no PB popular apresenta redução das marcas morfológicas, o que leva a considerar a influência das línguas africanas do grupo níger-congo atlântico em sua formação (GUY, 1981, p. 301-302). Estas línguas teriam transmitido essa característica de redução das categorias morfológicas aos crioulos de base portuguesa do atlântico e às variedades reestruturadas de português africano. Dessa forma, neste trabalho, é considerado o comportamento da regra de CV no PB pelo fato de ser uma língua resultante de contato linguístico com línguas africanas, principalmente.

Corroboram-se nesta dissertação os pressupostos de Guy (1981), de Baxter (1995) e Baxter e Lucchesi (1997) e de Holm (1988) por tratar da variação e da mudança linguística

¹ Ver com mais detalhes nas seções 1.2 e 1.3.

correlacionadas à sociohistória e à aquisição da linguagem. Ademais, por envolver falantes de uma variedade reestruturada de português, cujo substrato é o santome uma língua crioula.

Portanto, nesta pesquisa se propõe a complementar os estudos acerca da variedade de português falada na comunidade semi-isolada de Almoxarife em São Tomé, cujos membros são bilíngues em crioulo santome e português. Acredita-se que, através análise da CV-P4, pode-se verificar que o processo de aquisição da marcação de plural reflete a correlação existente entre as características tipológicas das línguas crioulas e da língua alvo (LA), o PE.

No português, a CV é realizada entre o sujeito e o verbo. Ambos possuem marcas de número e pessoa, mas o que permite a identificação do sujeito, quando este não está explícito foneticamente, é morfologia do verbo correspondente ao sujeito. Por conseguinte, há mais de uma forma de marcação da CV, como se pode verificar nos exemplos (1), (2) e (3) as possíveis realizações de concordância quando o verbo está em P4:

(1) *Nós eramos 5 numa canoa* [DULCEM1]

(2) *Vimos no memo caro* [DULCEM1]

(3) *Nós fica com fome* [FRANCM2]

Uma justificativa plausível para a ausência da marca de CV é a redundância. Observa-se em (1) que o sintagma nominal SN sujeito *nós* possui marcas de concordância com o seu sintagma verbal SV – *gostamos*. Em (2) o sujeito não está realizado foneticamente, todavia, é possível identificá-lo através do morfema - *mos*. No exemplo (3), há ausência de marca de CV, porém, mesmo assim, o sujeito da oração pode ser identificado.

A ausência da marca de CV em (3) se trata de mais um caso de omissão de redundância do que de falta de CV, uma vez que se pode identificar o sujeito, foneticamente expresso ou não, através de uma marca de CV, no caso do exemplo, a presença do pronome *nós* já dá a ideia de plural em toda a construção frasal.

Nas línguas crioulas, não há marcação de CV (cf. seção 1.1.2). Sendo a ausência de marcas flexionais uma das características das línguas reestruturadas, infere-se que no PA a ausência de CV é influenciada pelo substrato crioulo, envolvendo não somente fatores estruturais, mas também, extralinguísticos. Partindo desse entendimento, percebeu-se a necessidade de realizar um estudo de cunho variacionista sobre a CV-P4 na fala popular de São Tomé, pois se sabe que este ainda convive com uma forte influência das línguas crioulas (o santome, o angolano, e o caboverdeano), e que a sociedade, portanto, apresenta um quadro de multilinguismo (GONÇALVES, 2010, p. 15-19).

Esta dissertação se direciona a estudar o PST no que tange às diferentes realizações da CV-P4, investigando as seguintes questões:

i) No que se refere ao fenômeno da CV-P4, será que a gramática do PST se aproxima à do PB, considerando a sociohistória e o contato linguístico ocorrido na formação das variedades brasileira e são tomense do português?

ii) O comportamento morfossintático sincrônico da CV-P4 no PST é semelhante ou não ao PB?

As hipóteses gerais a serem testadas neste trabalho consistem em: (i) verificar o comportamento da variação da CV-P4 no PA, e (ii) se essa variação é motivada por fatores linguísticos e extralinguísticos que refletem características tipológicas, atuais e ancestrais, da L1 dos aprendentes de português.

Para finalidade de exposição do tema, a dissertação está dividida nos seguintes capítulos: no capítulo 1, apresenta-se o quadro linguístico de São Tomé e Príncipe, o percurso histórico das línguas crioulas e do português na ilha de São Tomé. No capítulo 1 também são discutidos aspectos relacionados ao comportamento dos verbos no santome e os motivos que levariam a CV a ser um traço não marcado no PA.

Ainda nesse mesmo capítulo são resenhados estudos já realizados acerca do tema da CV tanto para o PB quanto para o PST (LEMLE e NARO, 1977; Naro, 1981; NARO e SCHERRE, 1993; GUY, 1981; ALMEIDA, 2006; LUCCHESI, 2009; BRANDÃO e VIEIRA, 2012). Os pormenores desses trabalhos são discutidos a fim de verificar quais os fatores linguísticos e extralinguísticos que levam ao favorecimento e ao desfavorecimento da aplicação da regra de concordância nas variedades linguísticas estudadas, e que possam nortear a metodologia e os procedimentos analíticos da presente dissertação.

Outra discussão que se faz no capítulo 1 é acerca da aquisição de L1e L2, abordando também a forma como a Sociolinguística Variacionista, atualmente, tem estabelecido a relação entre o comportamento linguístico dos falantes de uma determinada comunidade de fala e a sua competência linguística enquanto usuários nativos de uma dada língua natural.

No capítulo 2, descreve-se o perfil sociohistórico da comunidade de Almojarife, listando alguns traços morfossintáticos que caracterizam a fala dos almojarifanos. São apresentados os procedimentos metodológicos adotados na investigação empírica, a exemplo da prática adotada na seleção dos dados, do programa estatístico utilizado e da amostra investigada, sendo especificadas as variáveis explanatórias utilizadas na investigação, bem como as hipóteses que norteiam a análise dos dados.

No capítulo 3, são apresentados os resultados acerca da variação analisada, que possibilitou a verificação do comportamento da variável dependente em relação aos diversos fatores das variáveis independentes. Os resultados obtidos são comparados com os alcançados em outras pesquisas e, é identificada uma tendência de mudança. Após o capítulo 3, seguem-se as considerações finais desta pesquisa, retomando os objetivos traçados, as hipóteses avaliadas e os resultados mais interessantes.

CAPÍTULO 1 - ENQUADRAMENTO TEÓRICO-DESCRIPTIVO DA CONCORDÂNCIA VERBAL DE PRIMEIRA PESSOA DO PLURAL

Neste capítulo tem-se a finalidade de mostrar o percurso histórico do quadro linguístico de São Tomé, bem como de apresentar o enquadramento teórico-descritivo do tema desta dissertação, a CV-P4. O capítulo começa com uma apreciação dos fatos históricos que contribuíram para a formação do atual quadro linguístico de São Tomé, bem como para a formação das línguas crioulas ali faladas (seção 1.1 e subseção 1.1.1). Em seguida, na subseção 1.1.2, considera como se deu a aquisição do português em São Tomé e Príncipe e como se processou sua transição de L2 para L1.

A seção 1.2 aborda os aspectos teóricos e descritivos do fenômeno linguístico. São comentados as matrizes metodológicas, as hipóteses e os resultados de algumas pesquisas variacionistas pioneiras sobre a variação na CV. São estudos que focam variedades do PB, sendo praticamente inexistentes estudos sobre essa variável em outras variedades do português.

Esses fatos remetem a outra questão: embora neste trabalho se trate da CV-P4, alguns dos trabalhos resenhados na seção 1.2 discorrem acerca da CV-P6 pelo fato de apresentarem metodologias e variáveis explanatórias de relevância para este trabalho.

Na seção 1.3 se trata da aquisição e da transmissão geracional, levando em conta aspectos da Gramática Universal e os fatos sociohistóricos. Como esta dissertação apresenta resultados do português como segunda língua, é essencial discutir como ocorre o processo dessa aquisição entre falantes adultos e como se dá a transmissão do português para outras gerações quando a aquisição ocorre pelo contato.

1.1 O PERCURSO HISTÓRICO DO QUADRO LINGUÍSTICO DE SÃO TOMÉ E PRÍNCIPE

Segundo dados sociolinguísticos (RGPH 1981; 1991; 2001), o português é falado por mais de 95% da população são-tomense, sendo que mais de 50% da população o falam como primeira língua, em detrimento do uso das línguas nativas, crioulas. Gonçalves (2010), apresentando os dados do Instituto Nacional de Estatística de São Tomé e Príncipe, comenta que, apesar de ser o menor país da África, com uma extensão territorial de cerca de 17 km² e um reduzido número de apenas 151.912 habitantes, as ilhas que formam o arquipélago de São Tomé e Príncipe possuem uma ampla diversidade linguística.

Segundo Hagemer (2009, p. 1), além da língua oficial, o português, são igualmente faladas três línguas crioulas autóctones, aparentadas: o santome – forro ou santomense - (língua de S. Tomé); o angolar (língua dos Angolares), ambos falados na ilha de S. Tomé, e o lung'ie (língua da ilha), falado na ilha do Príncipe, assim como o crioulo de Cabo Verde, o português dos Tongas e resquícios de línguas do grupo Bantu.

Após sua descoberta em meados do século XV, dois fatos históricos contribuíram para a formação do atual quadro linguístico de São Tomé: o povoamento e o ciclo do açúcar, no século XVI, e o ciclo do café e do cacau, nos séculos XIX e XX. Na fase do povoamento, a sociedade são tomense começou a ser formada a partir dos colonizadores portugueses e escravos africanos oriundos do antigo Reino de Benim, a região do delta do Níger (atual Nigéria), onde se falava e se fala o Edo – uma língua do grupo Edóide -, do Congo e de Angola, onde predominavam as línguas do grupo Bantu (HAGEMER, 2009, p. 2).

Com o desenvolvimento da produção de cana-de-açúcar para o comércio, houve necessidade de mão-de-obra, sendo esta trazida da região do Congo e de Angola. Se na fase do povoamento São Tomé esteve ligada à região do delta do Níger, esta fase da produção da cana sacarina fica associada à região onde predominavam as línguas bantu, mais precisamente o Kikongo e o Kimbundu, as quais se diferenciam tipologicamente do Edo do Reino de Benim (HAGEMER, 2009, p. 2-3).

Conforme Hagemer (2009, p.4), a proximidade entre realidades linguísticas distintas constituiu um fator primordial para a criação de uma língua própria dotada de léxico português e de estruturas gramaticais próximas das línguas africanas. Assim, do contato entre os europeus e os escravos e com a necessidade de comunicação, surgiu, então, um *pidgin*. Segundo esse autor, o *pidgin* foi desenvolvido no regime de habitação (fase de fundação da

colônia, antes do início da fase de produção agrícola), no qual o contato entre os europeus e os africanos era mais direto do que durante a plantação:

Especialmente no regime de habitação, em que o contacto entre europeus e africanos era mais directo do que no regime de plantação, havia desde cedo condições favoráveis a uma criouliização rápida. Cada povoador tinha direito a uma escrava por decreto régio, e as alusões à miscigenação são frequentes nos documentos antigos. As mulheres africanas e os filhos que nasciam dos casamentos e concubinatos com europeus foram oficialmente declarados livres a partir de 1515 e 1517, respectivamente, e rapidamente constituíram uma comunidade com reivindicações e poderes socio-económicos próprios. É plausível que esta comunidade de forros, escravos que recebiam a carta de alforria, com uma identidade própria, tenha estado na origem e consolidação da nova língua que se falava na ilha. O crioulo ter-se-á rapidamente difundido para as roças, no regime de plantação, tornando-se a língua-alvo dos escravos recém-chegados para efeitos de comunicação. (HAGEMEIJER, 2009, p. 4).

Apesar do período de produção da cana de açúcar ter sido bastante produtivo em termos linguísticos, o período da cultura do café e do cacau, em meados do século XIX, cunhou a história de São Tomé e Príncipe. Foi neste período que chegou às ilhas mão de obra de contratados provenientes das regiões de Angola, Cabo Verde, e Moçambique, incorporando suas línguas nativas ao quadro linguístico das ilhas (HAGEMEIJER, 2009, p.5). Assim, a partir do povoamento e da cultura açucareira, séculos XV e XVI, e da chegada dos contratados no período do café e do cacau, no séculos XIX e XX, fez desenvolver o quadro plurilinguístico de São Tomé e Príncipe.

1.1.1 As línguas crioulas em São Tomé

Segundo Hagemeyer (2009, p. 5-6) as línguas crioulas faladas em São Tomé possuem a mesma gênese. Usando como referência as obras de Günther (1973) e Ferraz (1974, 1979), Hagemeyer afirma que o contato linguístico decorrente do processo de povoamento de São Tomé resultou no surgimento duradouro de uma língua crioula de base lexical portuguesa que deu origem a outras quatro línguas. Para abordar melhor esse fato, nesta seção trata-se brevemente das quatro línguas crioulas faladas em São Tomé.

O *santome* ou São Tomense (santomense) é o crioulo dominante e de base portuguesa. Os falantes nativos o denominam de *lungwa santome* (língua de S. Tomé), *diôletu* (Dialeto) ou *fôlô* (Forro), que se originou da designação de escravos libertos – negros forros, visto que estes recebiam a carta de alforria. Segundo Gonçalves (2010, p. 11), o forro é a língua crioula

mais importante e com o maior número de falantes exatamente por esse motivo. Hagemeyer (2009, p. 6) afirma que os alforriados da primeira década do século XVI formaram um grupo com um estatuto social privilegiado, que chegou a se tornar proprietário de terras e de escravos e, devido a isso, a língua que eles falavam – o *forro* - se transformou na língua-alvo dos escravos recém-chegados.

O *angolar*, denominado pelos falantes nativos de *lungwa ngola*, é conhecido por ser a língua dos angolares, comunidade de descendentes de escravos que fugiram das roças durante o ciclo do açúcar. Conforme Hagemeyer (2009, p. 23), o *angolar* se diferencia do *santome* por sua base ser de origem bantu, e mais precisamente do kimbundu, e por apresentar uma parte do léxico de origem Edo, uma língua do grupo Edóide.

Além das línguas crioulas, São Tomé ainda apresenta em seu quadro linguístico o português dos Tongas, resultante do contato de descendentes de contratados, chamados de Tongas, falantes de kimbundu e umbundu com o português falado nas grandes roças de café e cacau da época colonial (HAGEMEIJER, 2009, p. 27-28). Além das línguas propriamente locais – *forro*, *angolar* e o português dos tongas – a ilha de São Tomé ainda abriga falantes do caboverdeano, apesar de não haver estudos acerca das características desta língua no arquipélago (GONÇALVES, 2010, p. 13).

Pelo fato de São Tomé apresentar um quadro amplamente plurilíngue, o português foi firmado como a língua oficial, após a independência (1975), a fim de unir os povos, de facilitar a comercialização, além da questão de prestígio nas classes sociais (GONÇALVES, 2010, p.14). Na atualidade, praticamente, toda a população fala português, deixando de lado o uso das línguas nativas.

Gonçalves (2010) afirma que, somente nos últimos anos, tem havido tentativas de preservação das línguas nacionais, a exemplo da realização de programas de rádio e televisão transmitidos em crioulo para que não sejam perdidos seus valores histórico e linguísticos. Na comunidade de Almojarife, a língua crioula falada é o *santome*.

1.1.2 O português em São Tomé e Príncipe

Como ficou exposto na subseção anterior, o português é a língua mais falada, em detrimento das línguas crioulas e demais línguas devido, principalmente, a fatores sociais, tais como a mobilidade social, o acesso generalizado ao ensino e aos meios de comunicação e a falta de políticas de manutenção das línguas crioulas. Hagemeyer (2009, p. 19), citando Afonso (2008) e Lorenzino (1996), afirma que, apesar de o PST seguir oficialmente a norma

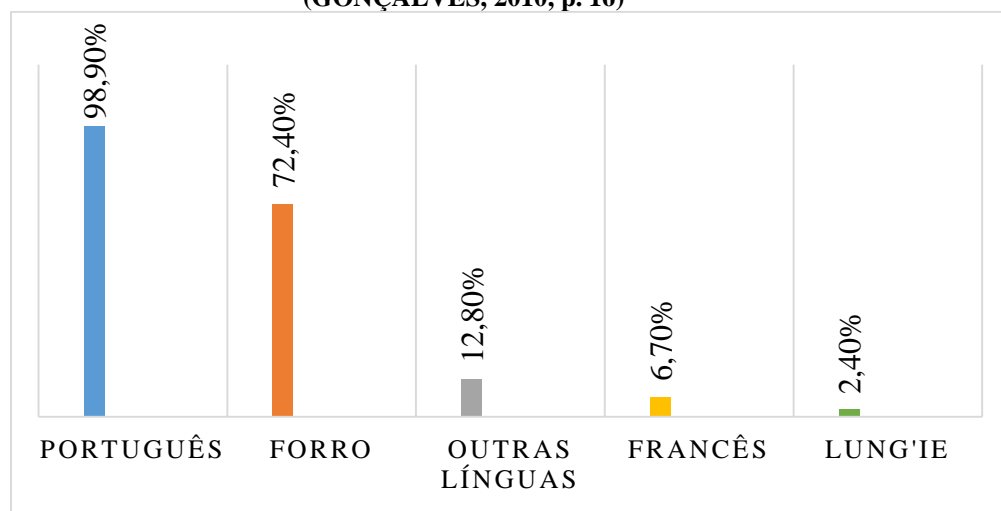
do PE, existem registros acerca da variação do português próximo dessa norma e do português com influência dos crioulos,

muitas vezes determinado por factores tais como o nível de escolaridade, nível económico e o ambiente de inserção social (urbano/rural). Esta variação reflecte o conflito entre a norma oficial e a prática local e um passado recente em que o português era L2 para a maioria dos habitantes das ilhas. (HAGEMEIJER, 2009, p. 19)

À medida que as línguas crioulas se tornavam presentes apenas em comunidades fechadas, no século XIX, os contratados que chegavam ao arquipélago iam adquirindo o português como L2 (Gonçalves, 2010, p. 15). Diante desse contexto, enquanto as línguas crioulas ficavam restritas às situações informais de comunicação e eram adquiridas em casa como L1, o português era aprendido na escola e praticado na sociedade como L2 ou L1. Gonçalves (2010, p. 16) afirma que o *input* (modelo-estímulo) disponível para a aquisição do português como L1 provém de variedades defectivas do português L2.

Dados do censo nacional de 2001, publicados em 2003 (RGPH 2003, p. 48) sobre as línguas faladas pela população mostram que o português é falado por 98,9% da população, seguido pelo forro com 72,4%. O grupo “outras línguas”, não especificado, registra uma percentagem de 12,8% de uso; o francês 6,7% e o lung’ie 2,4%. Gonçalves (2010) salienta que esses dados não permitem tirar conclusões se as referidas línguas foram adquiridas como L1 ou L2 e servem apenas para mostrar a distribuição linguística no espaço geográfico. O gráfico 1 apresenta os referidos percentuais.

Gráfico 1 – Distribuição da população por línguas faladas em São Tomé e Príncipe.
(GONÇALVES, 2010, p. 16)



Gonçalves (2010) também apresenta dados acerca da distribuição da população pelas línguas dominantes, português e forro, de acordo com a idade, evidenciando uma homogeneidade no uso do português nas diferentes faixas etárias. Contudo, há também uma diferença significativa entre os falantes de 20 anos, dos quais apenas 60% reportam falar crioulo, e os falantes com idade superior a 65 anos, dos quais 86,9% reportam um maior domínio do forro.

Esse fato revela que o português, inicialmente adquirido como L2, vem se transformando na L1 da população são-tomense. As estatísticas apresentadas também mostram que o processo de transmissão do português L2 para os falantes mais jovens, em competição com as línguas nativas e tornando-se L1 diante delas, leva ao surgimento de uma nova variedade de português (GONÇALVES, 2010, p. 20).

Cabe ressaltar que nesses casos, ainda ocorre um processo de erosão linguística (*language attrition*), que consiste no uso de uma nova língua em detrimento da competência da língua materna (WINFORD, 2003). Com as palavras de Batido (1992), Winford (2003, p. 258) descreve os cinco estágios do processo de erosão linguística:

- 1) A L1 é o único sistema linguístico de que o falante dispõe;
- 2) A L1 concorre com a L2, período de transição de monolinguismo a bilinguismo, em que o uso da L1 é restrito a ambientes informais, enquanto a L2 é a língua oficial;
- 3) Há uma permanente diglossia, a maioria dos falantes adotam a L2 em detrimento da L1, apontando para o seu apagamento e perda;
- 4) e 5) A proficiência da comunidade de fala na L1 original é cada vez mais limitada, culminando na completa “extinção” da L1 e consequente substituição pela L2. Todavia, os cinco estágios não são visíveis, pois a erosão linguística ocorre em tempos diferentes entre os diversos falantes.

Levando em consideração os estágios do processo de erosão linguística, pode-se inferir que São Tomé e Príncipe se encontram nos três últimos estágios, os quais são os responsáveis pela determinação da mudança linguística, ainda que haja a concorrência entre o forro e o português em algumas regiões. Gonçalves (2010) ainda acrescenta que

A emergência de uma variedade linguística envolve vários estádios de permanente interação e competição entre as gramáticas individuais dos falantes, nos quais são produzidos sistemas de regras que diferem do da língua-alvo. De acordo com Muysken (2008:141), o processo de erosão linguística tem consequências particularmente ao nível das categorias funcionais que tendem a ser muitas vezes

omitidas ou substituídas por outras equivalentes às da LM. (GONÇALVES, 2010, p. 22).

Acerca da influência das línguas crioulas no PST, mais precisamente no PA, sabe-se que o santome é o crioulo que mais influenciou a formação da gramática dessa variedade do português devido ao fato deste ser a língua crioula de maior alcance no país. O léxico do santome é, em grande parte, de origem portuguesa, mas suas características morfossintáticas estabelecem relações com as línguas africanas faladas pelos ancestrais escravos transplantados (HAGEMEIJER, 2009, p. 20). No caso concreto da estrutura verbal do santome, não há marcação de pessoa ou número no verbo, exemplos (4) e (5).

(4) *Ê saya kanwa pê matu.* (Santome)

Ele **puxar** canoa pôr mato. (PA)

“Ele **puxou** a canoa para o mato.” (PE) (HAGEMEIJER, 2009, p. 13)

(5) *Nós saya kanwa pê matu.*

Nós **puxar** canoa pôr mato. (PA)

“Nós **puxamos** a canoa para o mato.” (PE)

Para exemplificar melhor essa questão, segue-se um quadro 1 que compara a representação das pessoas e números em santome com o verbo ‘saber’, e as suas traduções em português:

Quadro 1. A expressão de pessoa e número com o verbo ‘saber’ em santome e português.

Santome	Português	
<i>N’ sebê</i>	Eu sei	OU eu soube
<i>Bo sebê</i>	Tu/Você sabe	OU tu soubeste
<i>E sebê</i>	Ele/Ela sabe	OU ele/ela soube
<i>Nõ sebê</i>	Nós sabe/sabemos	OU nós soubemos
<i>Nässe sebê</i>	Vocês sabe (m)	OU vocês souberam
<i>Inem sebê</i>	Eles/Elas sabe(m)	OU eles/elas souberam

Portanto, o verbo no santome não possui desinências número-pessoais nem flexão TMA, fatos que levam a crer que o sistema verbal do santome pode ter influenciado o do PA. No processo de aquisição do português por falantes do santome L1, teria ocorrido

transferência² do parâmetro da gramática do santomé para a L2 – o português. Em seguida, o mesmo sistema transitaria geracionalmente durante o processo de aquisição do português L2 com base nos dados linguísticos primários (DLP) defectivos transmitidos em cadeia, e que, conseqüentemente, determinaram uma reestruturação nos parâmetros do português L1. O processo envolvido é amplamente descrita por Winford (2003, p. 247-254).

A aquisição da marca de concordância constitui um fenômeno linguístico problemático para falantes do português L2, tanto para falantes instruídos como para os que não o são quando na língua nativa não há marcação de plural. O trabalho de Ao (2013) sobre a aquisição da morfologia verbal do português por falantes de chinês L1 confirma que esse fato é amplamente documentado na literatura científica.

Com base nos trabalhos de Lardiere 1998, Prévost & White 2000 e Hawkins 1997, 2003), Ao (2013) afirma que, no caso de falantes de chinês, aprendentes de português, isso ocorre porque enquanto a marcação da concordância é obrigatória na LA, é ausente na língua materna. (cf. AO, 2013, p. 2).

Ainda segundo Ao (2013) é consensual que os aprendentes de L2 tendem a apresentar variação morfológica a respeito de propriedades que são obrigatórias na LA e ausentes na língua nativa. Embora muitos estudos se debrucem sobre a aquisição dos traços abstratos, ainda há muito a explicar sobre as variações que ocorrem.

² Ver transferência funcional na seção 1.3.

1.2 PESQUISAS VARIACIONISTAS SOBRE A CONCORDÂNCIA VERBAL

Vários estudos acerca da concordância seguiram uma perspectiva comparada, analisando, principalmente, as variedades do PB e do PE (NARO, 1981; SCHERRE, 1988; NARO & SCHERRE, 1993, 2000, 2007), constatando que a marcação de concordância, seja verbal ou nominal, é um dos aspectos que evidenciam a distinção entre essas variedades do português.

No entanto, para as variedades africanas de português, pesquisas sobre a concordância nominal estão em estado mais avançado (MORENO e TUZINE, 1998; BAXTER 2004, 2009; FIGUEIREDO, 2008, 2009) do que para a CV, da qual poucos trabalhos foram realizados até agora (GONÇALVES, 2000; BRANDÃO e VIEIRA, 2012; LOPES e BAXTER, 2011). Esse fato torna evidente a necessidade da expansão de estudos que abordem as variedades africanas do português, bem como estudos comparativos entre essas variedades e o PB.

Os trabalhos sobre a CV aqui resenhados tratam não somente da CV-P4, mas também da CV-P6, por apresentarem variáveis explanatórias que podem ser pertinentes para a análise da P4, bem como por se tratarem de trabalhos pioneiros sobre o fenômeno. Vale salientar a presença de trabalhos da CV no PB devido à maior quantidade destes nessa variedade, bem como o fato de determinadas comunidades linguísticas possuírem características sociohistóricas que se aproximam do PST.

O trabalho precursor sobre a CV, cuja variável dependente foi P6, é o de Lemle e Naro (1977) “Competências básicas do português”, que analisou a fala de 20 informantes do Projeto Mobral³ no Rio de Janeiro: 9 mulheres e 11 homens; 6 informantes com mais de 40 anos e 14 com a faixa etária abaixo dos 30.

As entrevistas foram realizadas em situações diversas para verificar a influência do nível de formalidade sobre a CV. As variáveis controladas foram: circunstância da entrevista, definição ou indefinição do sujeito, posição do sujeito em relação ao verbo, diferenciação fônica entre as formas singular e plural do verbo (saliência fônica).

A variável *circunstância da entrevista* observou o possível efeito do contexto formal ou informal em que se deu o inquérito ao informante. A *definição ou indefinição do sujeito* avaliava se o sujeito da oração era identificado ou não. Em geral, os resultados dessa variável

³ O **Movimento Brasileiro de Alfabetização** - o **MOBRAL** surgiu como um prosseguimento das campanhas de alfabetização de adultos iniciadas com Lourenço Filho. Só que com um cunho ideológico totalmente diferenciado do que vinha sendo feito até então. Foi criado pela Lei número 5.379, de 15 de dezembro de 1967, propondo a alfabetização funcional de jovens e adultos, visando "*conduzir a pessoa humana (sic) a adquirir técnicas de leitura, escrita e cálculo como meio de integrá-la a sua comunidade, permitindo melhores condições de vida*"

estão associados aos da variável *posição do sujeito em relação ao verbo*. Em alguns estudos (ex. LUCCHESI, 2009), opta-se por unir essas duas variáveis em uma: *a posição e realização do sujeito*.

A variável posição e realização do sujeito avalia o potencial efeito da posição e realização do sujeito na aplicação da regra de CV. Essa variável tem sido contemplada em todas as análises variacionistas subsequentes sobre a variação CV no PB, tal é sua relevância que a constatação geral, segundo Naro e Scherre (2007), é que a presença de um sujeito imediatamente anteposto ao verbo favorece a CV, enquanto o sujeito posposto não influencia a CV. Contudo, outros estudos, como o de Lucchesi (2009) têm mostrado que a não realização fonética do sujeito (sujeito nulo) pode favorecer a CV através de uma lógica funcional, uma vez que a ausência do sujeito reforça a necessidade de indicação de pessoa e número do sujeito através da desinência verbal. (cf. LUCCHESI, 2009, p. 365-366).

O princípio da saliência fônica foi estabelecido por Lemle e Naro (1976) e, dada a sua relevância, passou a ser utilizado nos estudos de concordância que se sucederam. O conceito da saliência fônica envolve o potencial efeito sobre a CV do grau de oposição morfofonológica de singular/plural: quanto maior a distinção sonora entre a forma singular e plural de um nome, maior a probabilidade dele apresentar a marcação explícita de pluralidade, logo, quanto maior for a diferenciação morfofonológica entre as formas verbais singular/plural do verbo, é maior também a frequência de aplicação da regra de concordância.

Para “medir” o grau de saliência, são propostas diferentes níveis de saliência, que variam de acordo às desinências verbais a serem estudadas. O quadro (2) a seguir mostra os cinco níveis de saliência propostos por Naro *et ali* (2009):.

Quadro 2. Representação dos níveis de saliência fônica. (NARO *et all*, 1999, p. 203)

Nível	Oposição singular/plural
(1)	Ex.: <i>falava/falávamos; fosse/fôssemos</i>
(2)	Ex.: <i>fala/falamos; trouxe/trouxemos</i>
(3)	Ex.: <i>está/estamos; tem/temos</i>
(4)	Ex.: <i>comeu/comemos; partiu/partimos</i>
(5)	Ex.: <i>veio/viemos; é/somos.</i>

A maioria dos estudos de CV no PB, ao configurarem a saliência fônica como uma das variáveis, revela que os verbos mais salientes favorecem a concordância e os verbos menos

salientes desfavorecem a concordância. Logo, há uma propensão de os elementos mais salientes serem mais marcados, uma vez que, havendo um elemento morfofonologicamente mais perceptível, ocorre igualmente tendência para se aplicar a regra da pluralização.

Nos resultados do trabalho de Lemle e Naro (1977) foram especialmente significativas duas variáveis:

I – Saliência fônica: quanto menos saliente for a diferença entre singular e plural, mais provável será a falta de concordância;

II – A pós-posição do sujeito ao verbo desfavoreceu a CV é aquela em que o sujeito segue ao seu verbo, isto é, a menos saliente. Dessa forma, o elemento determinante da concordância segue ao elemento determinado, fazendo com que a falta de concordância seja menos óbvia. (LEMLE e NARO, 1977, p. 43-46). É importante lembrar que os resultados desse estudo se referem a dados de falantes com escolarização tardia, o que faz com que os resultados sejam particularmente notáveis. Em contrapartida, segundo Silvana Araújo (2014, p.280), nos estudos que abordam dados de falantes com pouca ou nenhuma escolarização, a saliência fônica tende a se mostrar como uma variável relevante ao passo que nos estudos com falantes cultos essa variável não se destaca, logo a escolarização e a saliência estão interligadas.

Outro trabalho relevante para os estudos sobre a concordância sujeito-verbo é o de Naro (1981). Levantando a hipótese inicial de que a não aplicação da CV é mais recorrente na fala de pessoas de classes sociais menos favorecidas, Naro (1981) investigou um *corpus* com 17 estudantes do MOBREAL como informantes: 9 homens com idade entre 16 e 20 anos em um grupo e 37 a 45 em outro grupo; 8 mulheres de 15 a 26 anos e de 43 a 45 anos.

Esse trabalho teve a P6 como variável dependente e apresentou como variáveis linguísticas explanatórias relevantes: a saliência fônica – nas sílabas mais acentuadas, a aplicação da CV foi maior -; e a posição e realização do sujeito – o sujeito que precede imediatamente o núcleo verbal e o sujeito não realizado favorecem a CV. O autor justifica este último fato explicando que se trata de um aspecto de caráter funcional: se o sujeito não é explícito, realizar a desinência de CV é fundamental para indicar a pluralidade.

No caso das variáveis extralinguísticas, Naro (1981) mostrou que a tendência foi a aplicação da regra de concordância por falantes do subúrbio, por mulheres e pelos mais velhos, ou seja, os fatores lugar de origem, sexo e idade se destacaram.

Os trabalhos de Anthony Naro e Marta Scherre também oferecem um leque de resultados muito relevantes como pontos de comparação em outras análises da CV. Os resultados de análises da variável P6 com dados do Projeto de Estudos do Uso da Língua

(PEUL) revelaram uma “tendência de marcas levarem a marcas e zeros levarem a zero ou, em termos mais gerais, uma tendência de formas gramaticais semelhantes ocorrerem juntas” (NARO e SCHERRE, 1993, p. 2).

Naro e Scherre (1993) também observaram o efeito condicionante da variável *paralelismo formal*, quando uma forma verbal marcada na oração anterior favorece a aplicação da CV. Por questões pragmáticas e funcionais, o morfema de pessoa e número também é bastante usado em verbos que constituem uma primeira referência, pois na primeira menção do sujeito, as marcas de referência tornam-se mais necessárias.

Dessa forma, a variável *paralelismo discursivo* também tem sido bastante observada nos estudos sobre a CV. Lucchesi (2009) explica esse fato a partir do princípio da coesão estrutural. Conforme esse princípio, quando há concorrência entre duas gramáticas, uma padrão e a outra não-padrão, se o falante começar a frase com a gramática padrão terá muito mais chances de concluí-la com a mesma gramática.

Os trabalhos pioneiros de Lemle e Naro (1977), Naro (1981) e Naro e Scherre (1993), além de fornecerem informações de grande relevância para os estudos acerca da CV, construíram uma matriz teórico-metodológica para o estudo da variação e mudança no PB, ao apontar a direção da mudança e sua origem.

Naro (1981), assumindo o conceito de deriva proposto por Sapir (1949 [1921]), defende a ideia de que as mudanças no PB estariam ligadas a uma deriva secular natural: uma tendência progressiva para a perda de flexões de número e pessoa. Essa mesma tendência estaria presente no PE e teria começado nos substantivos quando no final destes havia marcas menos salientes, a exemplo das formas nasalizadas. Dessa forma, a perda de flexão nos verbos teria começado também com a perda das marcas nasalizadas.

Um estudo que propôs uma explanação diferente daquela dos estudos de Naro e Scherre, foi o de Guy (1981). Nesse caso, a análise de dados de falantes do Rio de Janeiro levou em consideração também a sócio-história demográfica do Brasil. Guy afirmou que a variação da CV seria o reflexo de um processo de aquisição da morfologia de pessoa e número. Vinculada a essa afirmação, houve um debate acerca da causa da redução da morfologia em variedades populares e rurais do PB: a causa seria a deriva latina ou seria o contato linguístico presente na história do português no Brasil?

Guy (1981) estudou a CV-P6 em dados do *corpus* do Projeto *Competências Básicas* para analisar amostras de fala de vinte informantes (nove mulheres e onze homens) com idades entre 16 e 54 anos. Esses informantes faziam parte do MOBREAL, ou seja, estavam em fase de alfabetização, e eram socioeconomicamente desfavorecidos. Guy seguiu a matriz dos

estudos de Naro (1981), e chamou a atenção para os resultados das variáveis *realização e posição do sujeito, influência da fala do entrevistador, estilo da entrevista* (formal ou casual), *sexo do falante e faixa etária*.

Ao analisar os dados, Guy (1981) concluiu que quando o sujeito não foi realizado, os informantes tenderam a marcar o verbo para manter a oposição singular/plural, tratando-se de uma estratégia funcional para deixar claro o discurso. Quanto à posição do sujeito, o autor observou que o sujeito imediatamente anteposto ao verbo favorece a CV, e que quando há marcação de plural no sujeito, a probabilidade de realização da regra de CV aumenta. Os resultados também mostraram que quando o entrevistador marcava a CV, os falantes assim o faziam também. Guy justifica esse fato ao afirmar que se trata de uma estratégia de cópia ou de repetição do verbo da pergunta.

Em relação ao sexo do falante, Guy (1981) afirma que as mulheres tendem a preservar as normas de prestígio, ou seja, marcam mais a CV porque são mais sensíveis às pressões sociais. A aplicação da regra de concordância também foi maior na fala dos informantes mais jovens, os quais demonstraram um estilo mais monitorado na fala. Esses resultados apontaram para um quadro de mudança em curso com tendência para aquisição das marcas de concordância ao contrário do que tinha afirmado Naro (1981). Somente em estudos posteriores, Naro e Scherre (2000), ao fazerem novas análises sobre o PB, verificaram que há tendências tanto para a perda de flexão, como para aquisição e para a variação estável.

Situando os resultados do seu trabalho no contexto da história do Brasil em relação à presença maciça de africanos e sua importância na formação do país, Guy assevera que um padrão de variação de concordância condicionada pela saliência da oposição singular/plural é precisamente o que se esperaria encontrar em uma população descriolizante, que vai mudando do padrão proto-crioulo de não-concordância para o padrão *standard* de concordância. (GUY, 1981, p. 296).

Buscando responder à questão de quais seriam as motivações da variação da concordância nas suas variedades, outras pesquisas foram avançando. Alguns investigadores se firmaram sob a premissa da influência de substrato, a partir do que afirmou Guy (1981) sobre uma possível origem crioula do português popular do Brasil decorrente do contato de falantes do PE e falantes de línguas africanas no período colonial.

Esse fato que teria gerado uma reestruturação do português aprendido como L2 por transmissão linguística irregular (BAXTER, 1995; BAXTER E LUCCHESI, 1997; LUCCHESI, 2003), que engloba processos de mudanças provenientes do contato linguístico. Enquanto que outros não levaram em conta os aspectos sociohistóricos e seguiram a ideia de

que a ocorrência da variação é um fato interno inerente à evolução do português (NARO, 1981; SCHERRE, 1988; NARO E SCHERRE, 1993, 2000, 2007). Essas duas posições antagônicas produziram um saudável e acirrado debate entre os pesquisadores e motivaram o desenvolvimento de outros trabalhos com o objetivo de buscar evidências para uma ou outra posição. Porém, atualmente, é consensual a postulação da influência dos fatores sociohistóricos na formação do PB, os quais acentuaram a ocorrência de mudanças na estrutura dessa variedade linguística, levando a uma simplificação em sua gramática.

O estudo de Almeida (2006) investigou a CV nas 1ª, 2ª e 3ª pessoas do plural na comunidade de remanescentes de escravos São Miguel dos Pretos em Restinga Seca, no Rio Grande do Sul. Almeida analisou uma amostra de fala composta por 24 informantes, homens e mulheres, na faixa etária entre 15 e 90 anos, considerando a CV padrão *versus* a CV não-padrão das três das pessoas do plural, também observando a presença *versus* ausência das desinências (DNP's) de plural como um todo:

- DNP4 – nós plantamos
- DNP5 – vocês plantam
- DNP6 – eles plantam

As variáveis linguísticas consideradas foram:

- Referência do sujeito: verificar para qual das pessoas do plural o sujeito expresso ou apagado se refere;
- Conjugação verbal: para identificar a possível relação das conjugações com o emprego das desinências número-pessoais do plural;
- Posição do sujeito: seguindo os pressupostos de Lemle e Naro (1977), Almeida utilizou essa variável para observar até que ponto ela motiva a aplicação das DNP's em São Miguel dos Pretos. A análise da CV-P4 envolveu *sujeito anteposto separado do verbo por um constituinte de até duas sílabas*, *sujeito anteposto separado do verbo por vocábulos com mais de cinco letras* e *sujeito posposto*.
- Tipo de sujeito: sabendo que existe uma associação entre o uso da concordância e o tipo do sujeito na oração. Para a CV-P4, Almeida analisou a probabilidade de o sujeito expresso por pronome pessoal, demonstrativo ou relativo e o sujeito apagado motivarem a CV;
- Saliência fônica: também seguindo o postulado de Lemle e Naro (1977) acerca da influência da saliência fônica na regra de aplicação da CV, Almeida testou

essa variável utilizando a escala dos cinco níveis (NARO *et all*, 1999) para a primeira pessoa do plural;

- Tempo verbal: para verificar quais os tempos verbais que se mostram favoráveis ao emprego das desinências. Citando como referência as análises de Zilles *et all* (2000), Costa (1990) e Nina (1980), Almeida parte do princípio de que o imperfeito do indicativo desfavorece a presença da desinência de P4 (-*mos*), logo, os fatores foram: modo indicativo – presente sintético analítico, pretérito perfeito sintático e analítico, pretérito imperfeito sintético e analítico; uso do vamos como perífrase de futuro ou perífrase modal; modo subjuntivo – presente sintético, pretérito imperfeito sintético e analítico, futuro sintético; Infinitivo – infinitivo pessoal.

Os resultados desse estudo mostraram que os falantes da comunidade quilombola estão adquirindo a concordância padrão, pois os resultados por faixa etária mostraram que a geração mais nova apresentava o percentual de 40% de aplicação da regra da CV, enquanto a geração mais velha apresentava apenas 16%.

Acerca das desinências, houve 73% de emprego de DNP4 e 80% de DNP6. As variáveis linguísticas que se destacaram foram a saliência fônica e a posição do sujeito, isto é, as formas verbais mais salientes e o sujeito anteposto ao verbo favoreceram a aplicação da CV. Almeida também destacou o processo de aquisição das desinências número-pessoais, dando ênfase a um aumento de uso conforme a diminuição da idade, o que sugere um processo de mudança geracional. Para melhor exemplificação do uso das DNP's no PB, Lucchesi (2009) apresentam um quadro comparativo, no qual é possível observar os processos de variação que ocorrem no paradigma da flexão verbal nas duas normas do PB:

Quadro 3. Processos de variação no paradigma da flexão verbal no português brasileiro culto e no português popular do Brasil. (LUCCHESI, 2009, p. 333).

NORMA CULTA BRASILEIRA	NORMA POPULAR BRASILEIRA
eu trabalho	eu trabalho
você trabalha (~ tu trabalhas ~ tu trabalha)	você trabalha ~ tu trabalha
ele/ela trabalha	ele/ela trabalha
nós trabalhamos ~ a gente trabalha	nós trabalha ~ nós trabalh amo(s) a gente trabalha ~ a gente trabalh amo(s)
vocês trabalham (~ vocês trabalha)	vocês trabalha ~ vocês trabalh am
eles/elas trabalham (~eles/elas trabalha)	eles/elas trabalha ~ eles/elas trabalh am

O autor afirma que

O *continuum* que se verifica em relação à frequência de uso do morfema de número, desde a norma urbana culta, de um lado, até as comunidades rurais afrobrasileiras, no extremo oposto, passando pelo que se pode chamar de norma urbana semiculta e variedades populares urbanas e rurais do PB, constitui uma significativa evidência empírica de como o contato entre línguas afetou as diversas variedades do PB. As variedades populares que têm a sua formação mais fortemente marcada pelo contato entre línguas são aquelas que apresentam as menores frequências de uso do morfema de número, ao passo que as variedades urbanas culta e semiculta, que só indiretamente foram afetadas pelo contato em sua formação histórica, são aquelas que apresentam os maiores índices de aplicação da regra de concordância (LUCCHESI, 2009, p. 333-334).

Lucchesi (2009) afirma que para a P4, há um quadro de polarização sociolinguística semelhante ao que se apresenta junto a P6: alta frequência de aplicação da regra de concordância na norma culta *versus* a aplicação muito baixa da regra na norma popular. Tal fato é corroborado pela pouca existência de estudos sobre a P4 na fala culta, pois se trata de uma aplicação de CV quase categórica, enquanto a quantidade de estudos aumenta à proporção que se diminui o nível de escolarização.

Um estudo que se diferencia desse contexto de aplicação da CV condicionada diretamente pela escolarização é o de Almeida (2006), no qual a comunidade rural afro-brasileira analisada – São Miguel dos Pretos, Rio Grande do Sul – obteve uma frequência alta de aplicação da CV-P4, 73%. Almeida explica esse fato dizendo que os moradores dessa comunidade mantêm relações diretas com pessoas de outras comunidades, devido, principalmente, às necessidades de emprego saúde e educação. Logo, não se trata de falantes que vivem de forma mais isolada como é o caso das comunidades afro-brasileiras do interior da Bahia.

O trabalho de Lucchesi (2009) analisou a CV-P4 apenas com as formas verbais relacionadas ao sujeito representado pelo pronome *nós* em uma amostra de 24 entrevistas, composta por falantes analfabetos ou semianalfabetos, homens e mulheres, na faixa etária de 20 a mais de 60 anos de idade em quatro comunidades afro-brasileiras isoladas no interior da Bahia: Sapé, Rio de Contas, Helvécia e Cinzento.

O estudo contemplou a variação entre a marcação com os morfemas *-mos*, *-mo*, *-emo* e a sua ausência. Os resultados gerais revelaram um índice de 18% de aplicação da regra contra 82% de falta de concordância. As variáveis explanatórias selecionadas como mais significativas foram: realização e posição do pronome sujeito, paralelismo discursivo; e

saliência fônica. Para a realização e posição do sujeito, o contexto que mais favoreceu a regra de CV foi o sujeito não realizado, o que, segundo os autores, fez prevalecer a lógica funcional, em que a ausência do pronome reforça a necessidade da indicação de pessoa e número do sujeito através da desinência verbal (cf. LUCCHESI, 2009, p.366).

Os resultados da variável paralelismo discursivo confirmaram a hipótese, pois a frequência de aplicação da CV foi de 48% quando houve a presença de uma forma verbal marcada na oração anterior, enquanto a frequência diminuiu para 11% quando não houve marcação na oração anterior.

No que tange à saliência fônica, os resultados mostraram que quanto mais saliente for a diferença entre as formas plural e singular do verbo, maior a tendência para aplicação da CV-P4, assim como em estudos anteriores. O encaixamento social indicou um processo de mudança em curso em direção à implementação da regra de concordância. Esse fato foi atestado a partir da análise da variável faixa etária, que revelou uma ligeira inclinação no sentido do aumento do uso da regra da concordância verbal, conforme se avança dos falantes mais velhos para os falantes mais jovens.

A variável sexo mostrou que os homens tendem a empregar mais a regra de concordância verbal que as mulheres nas comunidades rurais afro-brasileiras estudadas. A variável estada fora da comunidade evidencia um quadro de mudança aquisicional de cima para baixo, pois os falantes que viveram pelo menos seis meses fora da comunidade são os que mais aplicam a regra de concordância.

Em um estudo comparativo, Brandão e Vieira (2012) compararam dados das variedades urbanas do PB e do PST e investigaram a concordância no âmbito do sintagma nominal SN e no também no âmbito da relação entre o verbo e o sujeito de P6. As análises foram desenvolvidas com base na fala de 34 indivíduos: 17 informantes brasileiros naturais de Nova Iguaçu, município da Região Metropolitana do Rio de Janeiro; 17 informantes de São Tomé que declararam ter o português como L1, moradores da capital do país.

Os resultados gerais mostraram que, para a CV-P6, em São Tomé, o índice de aplicação da concordância foi de 93,2%, enquanto que no PB foi registrado um percentual de 78,1%. As autoras elencaram em um quadro as variáveis linguísticas e as variáveis sociais selecionadas como estaticamente relevantes para o cancelamento da marca de plural.

Quadro 4. Variáveis atuantes para o cancelamento da marca de número no verbo por variedade.
(BRANDÃO e VIEIRA, 2012, p. 21)

PB	PST
Posição no nível oracional	Nível de escolaridade
Nível de escolaridade	Sexo
Saliência fônica	Saliência fônica
Animacidade do núcleo	Animacidade do núcleo
Posição do sujeito em relação ao verbo	Faixa etária
Faixa etária	Posição do sujeito em relação ao verbo

Brandão e Vieira (2012), ao analisarem os resultados, detectaram que a variável *nível de escolaridade* atuou de maneira semelhante em ambas as variedades do português. No PST, ficou evidente que quanto maior o nível de escolaridade dos informantes, menor foi o cancelamento da marca de plural: falantes menos escolarizados: PR .90 e frequência de 14,8% de não aplicação da regra, e falantes mais escolarizados: peso relativo (PR) .15 e 1,5% de não aplicação. No PB, os informantes de nível superior apresentaram PR .29 e 9,9% de cancelamento da marca de plural, enquanto os informantes do nível médio PR .69 e 32%, e os de nível fundamental PR .61 e 23,7% de frequência.

Em relação à saliência fônica, as duas variedades, revelaram uma tendência para maior cancelamento da marca de número em verbos que apresentam pouca diferenciação entre as formas singular e plural, como pode ser visto na tabela 1.

Tabela 1. Comparação dos níveis 1 e 4 de saliência fônica entre o PB e PST no cancelamento da CV-P6.
(BRANDÃO e VIEIRA, 2012, p. 23)

	PB		PST	
Nível 1 de saliência fônica (come-comem; fala-falam)	PR .60	23,9%	PR. 63	8,4%
Nível 4 de saliência fônica (veio-vieram; é-são)	PR. 26	10,2%	PR. 26	4,6%

A tabela 2 evidencia que tanto no PB quanto no PST, os resultados demonstraram que os sujeitos com o traço [-animado] favoreceram o cancelamento da marca de 3ª pessoa plural do verbo, enquanto os verbos cujos sujeitos apresentam o traço [+animado] se mostraram favoráveis à aplicação da CV.

Tabela 2. Comparação dos traços semânticos [+animado] e [-animado] do sujeito entre o PB e PST no cancelamento da CV-P6. (BRANDÃO e VIEIRA, 2012, p. 23)

		PB		PST	
[+ animado]	[muitos doentes] morriam nessa situação	PR .46	18,2%	PR .45	5,6%
[- animado]	[as dificuldades] não acaba	PR .70	33,5%	PR .71	13,7%

A posição do sujeito em relação ao verbo também foi selecionada nas duas variedades, mostrando que sujeitos pospostos favorecem o cancelamento da marca de plural de forma expressiva, apresentando PR .81 e um percentual 52,5% de frequência no PB, e PR .77 e 21,4% no PST. Enquanto o sujeito nulo favorece a CV, apresentando PR .69 com um percentual de cancelamento da regra de 31,3% no PB e PR .85 e 7,3% no PST. A variável faixa etária atuou de forma distintas no PB e no PST: no PST, os jovens aplicam mais a concordância; no PB, os informantes da faixa intermediária (36-55 anos), seguidos dos mais velhos, foram os que lideraram a CV, como mostra a tabela 3.

Tabela 3. Comparação da faixa etária entre o PB e PST no cancelamento da CV-P6. (BRANDÃO e VIEIRA, 2012, p. 24)

FAIXA ETÁRIA	PB		PST	
A-18-35 anos	PR .59	23,6%	PR .36	5,4%
B-36-55 anos	PR .43	17,8%	PR .64	6,7%
C-56-75 anos	PR .48	19,2%	PR .68	11%

A variável paralelismo oracional atuou somente no PB como favorecedora do cancelamento da CV, e nesse contexto, a variável sexo se mostrou relevante apenas no PST. Brandão e Vieira (2012) asseveram que

No plano linguístico, fica evidente a semelhança entre as duas variedades, não só no que respeita à seleção das variáveis, mas também às tendências verificadas para os fatores que as constituem. Ao que parece, as restrições estruturais que atuam no Português – que, por princípio, poderiam atuar no âmbito de qualquer língua – referem-se a componentes de ordem cognitivo-processual, que se revelam nos níveis sonoro (saliência fônica), sintático (saliência posicional, paralelismo) e semântico (traço de animacidade). No caso do PST e do PB, ao que tudo indica, é a atuação de determinados fatores externos que acelera ou detém a implementação dessas restrições linguísticas. No plano extralinguístico, por outro lado, os resultados da pesquisa demonstram que PB e PST se particularizam no que se refere à atuação dos grupos de fatores sociais. Embora haja convergências na seleção das variáveis extralinguísticas e, por vezes, nos índices dos fatores que as constituem, sua interpretação tem de levar em conta forçosamente a sócio-história de cada comunidade. (BRANDÃO e VIEIRA, 2012, p. 36).

Como foi possível notar, nessa seção foram resenhadas algumas pesquisas variacionistas acerca do fenômeno da CV no português com a finalidade de mostrar metodologias e variáveis explanatórias de relevância para este trabalho. As referidas variáveis foram saliência fônica, posição e realização do sujeito, paralelismo discursivo, sexo e faixa etária. Apesar de alguns dos estudos aqui apresentados não tratarem especificamente da CV-P4, acredita-se que os fatores linguísticos e não-linguísticos que atuam no fenômeno em outras pessoas verbais também sejam potenciais condicionantes da variação da concordância em P4, levando em consideração a interligação das pessoas verbais.

1.3 AQUISIÇÃO DE L2 E TRANSMISSÃO GERACIONAL

Conforme Chomsky (1981), o ser humano é provido de uma gramática inata, a qual está disponível desde o seu nascimento e vai se desenvolvendo de acordo com o seu crescimento. Para seu desenvolvimento, a criança toma por base a gramática dos adultos, sendo que não é uma imitação, mas sim a incorporação de um modelo para a língua materna. Chomsky ainda afirma que a criança possui um dispositivo de aquisição da linguagem (DAL), que é ativado e se processa a partir da produção sentenças, o *input*, dos adultos, que resulta na gramática que a criança está contextualizada.

De acordo com essa proposta, toda criança possui uma Gramática Universal (GU), inata, um conjunto de princípios que podem ser realizados através de parâmetros conforme a língua à qual a criança é exposta, sendo que a criança modifica esta gramática de acordo com a gramática de sua língua, deixando só o que necessário para o uso e aprendizagem da mesma e descartando o restante. Para Chomsky (1981), a criança adquire e desenvolve sua gramática interna a partir de evidências e estímulos contidos nos DLP e na GU inata, sendo que os estímulos desencadeiam os parâmetros da GU, definindo a gramática da língua a ser usada.

Para a Teoria de Princípios e Parâmetros, a GU é o estágio inicial da aquisição da linguagem. Nesse sentido, a linguagem é formada por dois conjuntos de elementos: os princípios universais comuns a todas as línguas e, os parâmetros particulares ainda não traçados pela experiência do indivíduo com a língua materna. É durante o processo de aquisição que a GU retira informações da língua ambiente do falante e formata seus parâmetros⁴. (KENEDY, 2013, p. 97)

Como fora dito, o ambiente linguístico ao qual a criança é exposta ativa os parâmetros, permitindo o desenvolvimento da linguagem, assim toda criança é capaz de aprender sua língua materna nos primeiros anos de vida - entre 1 e 4 anos de idade. No entanto, por vezes, o *input* que a criança recebe pode ser incompleto, possuindo hesitações e interrupções, incluindo frases agramaticais e formas parciais tanto quanto frases gramaticais. Embora isso possa ocorrer, a criança aprende a língua materna e consegue se comunicar em um curto tempo de vida devido à base biológica da GU e ao DAL. (WHITE, 1989; KLEIN, 1986).

⁴ Como a CV é um parâmetro, não é comum à todas as línguas, logo, ela é variável. Isto é, algumas línguas marcam-na como positiva e outras como negativa. O português marca a CV como positiva, já nas línguas crioulas, por não possuírem este parâmetro, a marcação é negativa. Isso faz com que o adulto falante nativo de crioulo encontre dificuldades em adquirir os traços da CV existentes no português durante o processo de aquisição.

Na aquisição de uma L2 por um aprendente adulto, segue-se o mesmo mecanismo da aquisição da língua materna: extrair sentido de dados linguísticos, de produzir um sistema que explique tais dados e que lhe permita compreender e produzir estruturas na L2. O aprendente adulto de L2 utiliza as mesmas ferramentas que a criança na língua materna, isto é, os princípios e parâmetros da GU.

Porém, ressalta-se duas questões: primeiramente, o fato de que o problema da aquisição de L2 possa talvez ser resolvido sem o uso da Gramática Universal, mas através do uso de conhecimentos advindos da língua materna do aprendiz. Em segundo plano, o problema pode não ser resolvido, pois a maioria dos aprendizes de L2 não conseguem chegar a níveis avançados de domínio da LA e dificilmente chegam a desenvolver um sistema linguístico que seja qualitativamente superior ao *input* recebido. (WHITE, 1989, p.38).

Uma questão em que há divergência no âmbito da aquisição de uma L2 é o acesso à GU, isto é, até que ponto o aprendente adulto teria acesso aos parâmetros e princípios da GU e se esse acesso seria pleno, parcial, dual ou não seria possível ter este acesso. São postuladas as seguintes hipóteses:

1. Acesso pleno: neste o processo de aquisição de L2 seria igual ao processo de aquisição da língua materna, no qual os falantes têm acesso total à GU. Assim, a aquisição da L2 seria resultante da GU e dos dados da L2 e a língua materna não interferiria neste processo;

2. Acesso parcial: para essa perspectiva, os aprendentes de L2 teriam acesso aos parâmetros da GU que são operantes em sua língua materna, havendo a possibilidade de reativar os parâmetros relevantes para L2 através do ensino formal;

3. Nenhum acesso: segundo essa posição, os parâmetros da GU não estariam mais disponíveis aos aprendentes adultos de L2. Dessa forma, seriam usadas estratégias gerais de aprendizagem, o que provocaria um elevado grau de variação na aquisição da L2, levando os adultos a não conseguir atingir a LA;

4. Acesso dual: neste os aprendentes adultos conseguiriam fazer uso tanto da GU quanto das estratégias de aprendizagem. Mas, o uso dessas estratégias bloqueariam o acesso à GU, fazendo com que os aprendentes não atingissem a LA e chegassem a produzir sentenças não aceitáveis pela GU. (ELLIS, 1997, p. 69)

Em outras palavras, no plano da Teoria da Gramática, a discussão gira em torno de qual seria o nível de acesso dos falantes adultos aos dispositivos da GU no processo de aquisição de L2. Caso a resposta seja negativa, os falantes só teriam acesso aos dispositivos gramaticais e aos parâmetros fixados na sua língua nativa, fazendo ocorrer uma reestruturação

gramatical através de processos como **transferência funcional** e de **relexificação** (HOLM, 1988; LEFEBVRE, 1998).

Entende-se por **transferência funcional** a incorporação de material morfológico da língua de substrato ou o uso de material lexical da língua superstrato em funções morfológicas presentes – de forma plena ou parcial – na língua de substrato. Essa transferência ocorre através do processo de gramaticalização, que se trata da ampliação das funções gramaticais da língua de superstrato incorporados na L2 emergente e da utilização de itens lexicais para expressar relações gramaticais⁵.

A **relexificação** é a substituição total ou parcial do léxico de uma língua pelo de outra, isto é, a relexificação é um processo mental que cria entradas lexicais copiando entradas lexicais de um léxico pré-existente, substituindo suas representações fonológicas por representações derivadas de outra língua. É importante ressaltar que a relexificação é direcionada pela semântica, logo, para que ela ocorra, é necessário que as entradas lexicais da L1 e da L2 tenham algo em comum semanticamente⁶.

Sabendo que o *corpus* do presente trabalho é composto por informantes adultos bilíngues em português e crioulo, é relevante verificar os fatores que condicionam a aquisição de L2, bem como a sua transmissão. Nesse sentido, atrela-se a questão do acesso à GU e os fatores da transmissão linguística normal e a transmissão linguística irregular para compreender o processo de aquisição do português em Almojarife.

Na transmissão normal, a geração de falantes mais velhos fornecem DLP a partir do seu desempenho na língua nativa, isto é, os parâmetros da gramática das crianças são baseados no modelo da gramática dos pais. De maneira oposta, no processo de crioulição,

⁵ Exemplos do processo de transferência funcional e gramaticalização são os demonstrativos *kel* (derivado do português *aquele*) e *se* (do português *esse*) que podem funcionar como artigo definido nos crioulos de Cabo Verde e de São Tomé (LUCHESE, 1993)

⁶ A hipótese da relexificação pressupõe que as entradas lexicais dos crioulos (e dos pidgins) terão as propriedades semânticas e sintáticas do substrato e uma representação fonológica derivada da língua de superstrato, também chamada de língua lexificadora. Veja-se o exemplo do crioulo haitiano:

- (6) a) *Nèg la vòle machin (nan) an* (haitiano)
- b) *Súnù ò fin mótò (ó) ó* (fon)
- c) *homem Det roubar carro Det Det*
- d) *O cara roubou um/o carro.*

É o caso de concordância no interior de AGRP-S no qual os itens lexicais são derivados do francês, mas a gramática e a semântica são africanas (fon), inclusive o determinante de frase. Ainda de acordo com a relexificação, os aprendizes de uma L2 se apropriam de um novo vocabulário associando palavras e frases de suas L1 a palavras e frases de L2, uma a uma, frequentemente ignorando os contextos linguísticos e pragmáticos em que são originariamente usadas. Essa associação facilita a aprendizagem de L2 em situações de pouco acesso à LA. (LEFEBVRE, 1997, p. 185)

os DPL para a aquisição da língua materna são fornecidos por adultos que adquiriram a L2 pelo contato, e contêm material linguístico que diverge daquele da LA do processo de aquisição de L2.

A partir disso, uma questão é levantada:

Quem são os principais agentes desse processo? Os adultos, que recorrem ao repertório gramatical de sua(s) língua(s) nativa(s) para reestruturar gramaticalmente o código de comunicação emergencial? Ou as crianças, que criam uma língua nova a partir de um *input* muito precário em função dos dispositivos inatos da Gramática Universal que atuam no processo de desenvolvimento da língua materna? (LUCCHESI E BAXTER 2009, p. 103)

A partir do questionamento de Lucchesi e Baxter, citado anteriormente e da observação empírica do *corpus* da comunidade de Almocharife, acredita-se que tanto os falantes mais velhos quanto os mais jovens sejam agentes atuantes centrais no processo de formação de uma L2.

Supõe-se que os dados linguísticos dos aprendentes adultos mostram que estes teriam um acesso parcial à GU, o que leva os falantes a preencherem as lacunas da L2 com parâmetros da língua materna. E, por sua vez, as gerações de aprendentes por vir aprendem uma nova língua, originada de um *input* escasso, com a possibilidade de corrigir as lacunas com a instrução formal.

Diferentemente do processo normal de aquisição da linguagem, proposto por Chomsky, há o caso de aquisição de L2 por falantes adultos em situação de aprendizagem precária e/ou forçada, no qual a transmissão geracional ocorre em condições adversas de aprendizado. É o que se conhece por *transmissão linguística irregular*.

Transmissão linguística irregular é o termo usado para designar o produto dos processos históricos de contato maciço entre línguas de tipologias diferentes (BAXTER, 1992, 1995; BAXTER E LUCCHESI, 1997; LUCCHESI, 2003; 2009). Levando em conta que esse contato linguístico se deu em uma situação de dominação, em que a língua do dominador, a língua de superstrato ou LA, era imposta para que os falantes das outras línguas pudessem adquiri-la, mas não haviam condições favoráveis ao aprendizado.⁷

⁷ Nos casos de contato entre línguas podem ocorrer ou situações que levam à formação de um crioulo radical, como o crioulo santome, ou, ainda, um processo de transmissão linguística irregular do tipo mais leve, como deve ter ocorrido nas variedades populares do português do Brasil, em que não se originou outra identidade linguística, mas uma variedade da língua alvo. Lucchesi e Baxter (2009, p. 102) afirmam que: “Dessa forma, o processo de transmissão linguística irregular pode conduzir à formação de uma língua historicamente nova, denominada **língua crioula**, ou à simples formação de uma nova variedade histórica da língua de superstrato, que não deixa de apresentar processos de variação e mudança induzidos pelo contato entre línguas.”

Diante desse contexto, se formaram variedades de L2 com particularidades distintas da LA, dado o fato do contexto sócio-histórico em que originaram. E, assim, novos modelos de aquisição da língua materna foram transmitidos à proporção em que as línguas nativas foram sendo deixadas.

Em suma, o processo de transmissão linguística irregular se diferencia do processo de transmissão geracional normal das línguas humanas, pois neste último as crianças dispõem de dados linguísticos completos, sem lacunas e reanálises. Portanto, a partir do processo de transmissão linguística irregular pode se formar uma nova língua, ou crioula ou uma nova variedade da LA.

Os DLP disponíveis para as crianças que nascem nessa situação, para desenvolvimento de sua língua materna, advém, quase categoricamente, da L2 dos falantes mais velhos, sendo que nesta L2 há reanálises dos mecanismos gramaticais. De acordo com Lucchesi e Baxter, (2009), as variedades de segunda língua que se formam em condições de contato linguístico podem ser mais ou menos defectivas consoantes as especificidades de cada contexto histórico. Dessa forma, a nativização⁸ da língua ocorre de maneira irregular, visto que os DLP disponíveis para as crianças nascidas nesse contexto são de versões com reanálises e lacunas de L2, desenvolvidas entre os adultos falantes de outras línguas.

Os falantes adultos, ao aprenderem uma L2 em condições precárias, formam uma nova variedade da LA que apresenta uma grande reestruturação gramatical, visto que somente os elementos essenciais para a comunicação são fixados na aprendizagem defectiva. Segundo Lucchesi e Baxter (2009), a redução na estrutura gramatical da língua deve-se a três fatores: I – à dificuldade de acesso dos falantes das outras línguas aos modelos da LA; II – como a maioria dos falantes são adultos, os dispositivos inatos que agem naturalmente na aquisição da língua materna já não estão disponíveis; III – não há uma norma ideal para a orientação da aprendizagem no processo de aquisição da LA. (cf. BAXTER E LUCCHESI, 2009, p. 102)

Semelhantemente ao PB, a aquisição do PST se deu justamente numa situação de contato maciço entre línguas distintas, dado o seu contexto sócio-histórico. Esse fato fez com que o PST se configurasse em uma nova variedade do PE, que era a LA de seus falantes. Resta saber qual o nível de distinção entre essas variedades, o que só é possível através de pesquisas empíricas mais profundas.

⁸ O termo nativização é utilizado para definir o processo segundo o qual uma anterior segunda língua é adotada como a primeira língua - e língua nativa - de uma comunidade, conforme Winford 2003, p. 244.

Uma assertiva relevante sobre o PB pode ser usada para o PST, dada a sua analogia no processo sociohistórico, foi feita por Lopes e Baxter (2009) afirmando que devido aos processos variados de aquisição do português que ocorreram na história do país, a aquisição de Português L1 se deu com dados divergentes ou insuficientes oriundos de contatos com falantes que aprenderam essa língua como L2. Nessa perspectiva, Lightfoot (1989) aponta que, embora seja controlado, o organismo linguístico é desenvolvido através de estímulos ambientais no processo de aquisição.

Se a característica marcante do contato é a redução da estrutura gramatical, Lucchesi e Baxter (2009) apontam que as estruturas mais abstratas que não possuem funcionalidade no ato da comunicação são dispensados, assim, as marcações morfológicas de caso, número, pessoa, tempo, modo e aspecto, e as relações sintáticas de regência, concordância, coordenação e subordinação são dispensados. Dessa forma a aquisição da marca de CV seria precária, uma vez que a concordância obrigatória na LA, é ausente na língua nativa (cf. subseção 1.1.2 deste trabalho).

1.4 SÍNTESE

Neste capítulo foi apresentado, em primeiro lugar, o percurso histórico das línguas crioulas, bem como do português em São Tomé buscando compreender a configuração dos processos de aquisição linguística e transmissão geracional no percurso histórico dessas línguas na ilha.

Em São Tomé a língua portuguesa foi adquirida num contexto de L2 e foi transformado na L1 da população em detrimento das línguas crioulas. Como consequência dos processos históricos, o que resulta desse processo é a variação sociolinguística e divergências no PST quando comparado com o PE.

Foram apresentadas pesquisas já realizadas sobre o tema da CV PB e no PST, não somente de P4, mas também de P6, com a finalidade observar os fatores linguísticos e extralinguísticos que favoreceram a aplicação da regra de concordância. Após a explanação das pesquisas foi possível identificar que há variáveis relevantes comuns aos estudos sobre a CV, as quais também serão utilizadas para a análise dos dados da presente pesquisa.

Por fim, discutiu-se como se dá o processo de aquisição de L2 e transmissão geracional junto aos princípios e parâmetros da GU em consonância com o processo de transmissão linguística irregular nas situações de contato maciço entre línguas. Tendo em vista que o português como L1 da maioria dos falantes nas ilhas foi adquirido a partir da L2 defectiva, sendo então, a L1 uma variedade reestruturada do português.

CAPÍTULO 2 - METODOLOGIA

Neste capítulo, em primeira instância, é apresentada a comunidade de Almojarife, em São Tomé, da qual foram extraídos os dados que compõem o *corpus* deste trabalho. É apresentada a socio-história da comunidade na seção 2.1, bem como seus dados sociolinguísticos com a caracterização do português falado em Almojarife, apresentando também alguns aspectos sociais dos informantes do *corpus*, na subseção 2.1.1.

A subseção 2.1.2 mostra a origem dos dados e a estrutura do *corpus* do trabalho, seguida da seção 2.2, na qual é brevemente apresentado o enquadramento teórico deste trabalho. Na subseção 2.2.1 são mostrados os aspectos metodológicos utilizados nesta dissertação, bem como a chave de codificação, o programa computacional utilizado e quantificação dos dados. Por fim, em 2.3 são apresentadas as hipóteses deste trabalho baseadas nos aspectos teóricos da Sociolinguística Quantitativa (LABOV, 2008 [1972]) e nas leituras de diversos trabalhos acerca do fenômeno da CV.

2.1 A COMUNIDADE DE FALA EM FOCO

Em 1644, quando os portugueses reconquistaram São Tomé dos holandeses, encontraram a ilha dividida em duas partes, numa situação de contato linguístico entre os membros da população no interior da ilha: escravos africanos e seus descendentes, a maioria falante do santome, outros bilíngues em forro e português L2 ou português L1 nativizado, e nos litorais: colonizadores, falantes do PE, e seus descendentes bilíngues em santome e português L1.

Henriques (1989 *apud* FIGUEIREDO, 2010, p. 85) diz que os africanos deixavam de estar inseridos no espaço linguístico-cultural dos colonos ao se refugiarem no mato fugindo do trabalho explorador nas lavouras de cana-de-açúcar, ao passo que a economia desse setor caía e afastava uma nova entrada de colonos na ilha, o que os separava linguisticamente cada vez mais da população das áreas rurais.

Diante desse contexto, a economia só viria a melhorar com a inserção das culturas do café e do cacau (1800 e 1822, respectivamente). Porém, como estavam vivendo em uma situação de autossuficiência no interior da ilha, a maior parte dos libertos e descendentes de africanos não tinha interesse em trabalhar para os colonos, fazendo com que a mão-de-obra tivesse que ser contratada vindo de Angola e Moçambique no final do século XIX, e de Cabo Verde em meados do século XX.

Mesmo após a abolição da escravatura, datada de 25 de fevereiro de 1869, os trabalhadores das roças eram obrigados a trabalhar em condições desumanas. Nas palavras de Ribeiro (1875), Figueiredo (2010) relata que essa situação levou os libertos a não buscarem empregos nas roças, se fixando em comunidades isoladas. Assim como no Brasil, a maioria deles procurava sobrevivência instalando-se nas periferias urbanas, enquanto outros permaneciam em comunidades semi-isoladas como ocorreu com os forros em Almojarife.

Segundo Figueiredo (2010), não há uma documentação que comprove com exatidão a origem da comunidade de Almojarife, porém o fato de ocorrer bilinguismo entre o santome e o português reestruturado deixa indícios que ela não tenha se originado na ocasião da desativação dos engenhos de açúcar, o que determinou a fixação de vários engenhos comunitários no interior da ilha.

O grupo de forros que se estabeleceu na Baía de Almojarife seria constituído por libertos são-tomenses que buscavam se fixar longe das roças para não serem forçados a trabalhar nelas. O sustento da comunidade se deu através da agricultura de pequeno porte e da pesca artesanal.



Figura 1. Índice demográfico de Santana e Almozarife (Tenreiro, 1961, p.164).

A Baía de Almozarife situa-se ao leste da ilha de São Tomé, pertencendo ao Distrito de Cantagalo, cuja capital é Santana. Dados do censo de 1960 mostram que o distrito contava com cerca de 13.258 habitantes, a capital com 7.702, e a comunidade de Almozarife possuía cerca de 150 forros. (TERNEIRO 1961, p. 164 *apud* FIGUEIREDO, 2010, p. 87).

Além do receio dos membros da comunidade em regressar ao trabalho nas roças, questões de cunho político e religioso e o difícil acesso aos meios de comunicação também fizeram com que os almozarifanos permanecessem por muito tempo com pouco contato com falantes do PE.

O português reestruturado falado pelos núcleos urbanos e pelas comunidades rurais mais isoladas, sobretudo nas gerações mais novas, começou a nivelar-se a partir da década de 1970 com a construção e pavimentação de estradas. Nessa fase, o contato de falantes bilíngues de Almozarife com o PE permitiu a inserção de empréstimos lexicais e estruturais da LA, colocando as duas gramáticas em competição e originando o português falado atualmente na comunidade (FIGUEIREDO, 2010, p. 88)

Após a independência de São Tomé e Príncipe, em 1975, seguiu-se um período marcado por conflitos internos e com a economia fragilizada. Devido a isso, as comunidades rurais ficaram abandonadas e cada vez mais isoladas, uma vez que ocorreu a desarticulação dos sistemas educacional, rodoviário e de comunicação (cf. FIGUEIREDO, 2010, p. 88-89).

Um dos aspectos no qual Almozarife sofreu com essa desarticulação foi no que tange à escolarização, pois se no Distrito de Cantagalo cerca de $\frac{1}{4}$ dos jovens encontra-se em idade escolar, nas comunidades mais afastadas a situação se agrava. E com a desarticulação do

sistema educacional, separando o PE das variedades populares, fez com que apenas a classe alta da sociedade utilizasse a norma padrão do PE, apesar de 98, 9% da população usar e compreender português (Figueiredo, 2010, p.89).

2.1.1 Características do português de Almojarife

O que contribuiu muito para a situação de semi-isolamento de Almojarife antes da independência de São Tomé foi o difícil acesso à capital do distrito, Santana. Como só havia uma estrada de terra batida, e o percurso tinha que ser feito a pé, apenas comerciantes faziam o trajeto. Após a independência, o isolamento aumentou, pois os colonos deixaram a ilha, as rodovias se tornaram intransitáveis e a mobilidade com o transporte público deixou de existir (FIGUEIREDO, 2010, p. 91)

No *corpus* deste trabalho, há menção da dificuldade de contato dos membros da comunidade com os moradores de Santana, como relata o informante [OSVALH1] em seu diálogo com o entrevistador:

DOC: *Você custuma i a Santana?*

INF: *Vô umas vezes devido quê zê essa prigiça de i a pé e volta a pé.*

DOC: *Hum, mas você vai assim em Santana mais pra fazê o quê?*

INF: *Quê zê ê vô mais a Santana assisti programa.*

DOC: *Hum.*

INF: *Só!*

Conforme Figueiredo (2010, p.92) o afastamento linguístico do PE foi mantido, fazendo com que as características defectivas do português L2 fossem conservadas. Tal é esse fato que os falantes mais velhos (acima de 65 anos) afirmaram ter o santome como L1, mesmo tendo aprendido o português L1 através do contato português L1/português L2 da geração anterior a deles.

Os falantes da faixa etária intermediária (41-65 anos) possuem o santome como L1, somente uma informante, [MANEVEM2], aprendeu esta língua e o português paralelamente. Alguns dos falantes dessa faixa etária trabalharam nas roças, levando o português de Almojarife (PA) a manter contato com o português dos “contratados”, impregnado de características das línguas bantu (FIGUEIREDO, 2010, p.92).

Portanto, o PA manteve aspectos gramaticais divergentes em relação à LA, sendo que esses aspectos constituíram o modelo disponível para o processo de nativização do PA como

L1, adquirido pelos mais jovens (falantes de 20 a 40 anos) simultaneamente ao santome. Ressalta-se que o português era a língua de prestígio e o santome estigmatizado na aquisição bilíngue, como se vê no relato da informante [ANTOM1]:

- DOC: *Qual qual foi a língua que você aprendeu primeiro?*
 INF: *Português.*
 DOC: *Não aprendeu dialecto em casa não?*
 INF: *Ê també aprendi. [Põe lá pra mim, Sã!]*
 DOC: *Sua vó falava em casa com ocê como?*
 INF: *Falava dialecto também.*
 DOC: *Falava mais dialecto ou português em casa, sua avó?*
 INF: *Hum, mas assim quan gen fala dialecto assim senhora não gostava.*
 DOC: *Ahã tinha que falar português?*
 INF: *Hum.*
 DOC: *A... a criança que fala dialecto com mais velhos é falta de respeito?*
 INF: *Não.*
 DOC: *Hum, pode falar?*
 INF: *Pode falá.*
 DOC: *Mas por que que sua avó não gostava?*
 INF: *Minha avó diz gen era muita criança ainda quê pa tá usa dialecto e assim quem gen pa um meio de pissoa vai falar assim mesmo.*
 DOC: *Hum.*
 INF: *Calha feio.*

Seguem-se algumas diferenças entre o PA reestruturado e o PE, tanto no âmbito fonológico quanto no nível morfossintáticos: abaixamento na realização das vogais altas, não realização de algumas preposições, variação na CN e na CV, ocorrência típicas do processo de transmissão linguística irregular (cf. FIGUEIREDO, 2010, p. 93), que no caso de Almojarife demonstra que há duas gramáticas em competição: a crioula de substrato santome e a portuguesa.

Através da observação do *corpus* da comunidade, foi possível verificar que os fenômenos que ocorrem no PA são mais incidentes entre os falantes idosos, levando a crer que a variação seria mais elevada quando ocorreu a formação da comunidade. Figueiredo (2010, p. 94) também observa essa questão e afirma que no PA é possível visualizar estruturas variáveis que decorrem da aquisição defectiva da gramática do PE, tais como⁹:

⁹ A lista de exemplos dos fenômenos que ocorrem no PA foi retirada do estudo de Figueiredo (2010, p. 94-97), que analisou a variação e a aquisição da CN em Almojarife, fazendo uso do mesmo *corpus* da presente pesquisa.

- Variação na colocação do possessivo:
PA: *E amigo meu tinha uma mulher* [OSVALH1]
“E o meu amigo tinha uma mulher”
- Concordância variável de número entre os elementos do SN
PA: *já trabalhê esses lugar tudo* [LUISH2]
“já trabalhei nesses lugares todos”
- Variação na flexão de gênero dos nomes referentes a seres animados, com recurso ao valor referencial do sexo usando nomes adjetivados:
PA: *Dois filho mulher qu'ê tem* [FRANCM2]
“As duas filhas que eu tenho”
- Concordância variável de gênero entre os elementos do SN:
PA: *faço trabalhos caseira até tarde* [CARMOM1]
“faço os trabalhos caseiros até tarde”
- Uso de formas do presente para expressar ações e estados passados:
PA: *É os meus pai que me deu essa ideia.* [OSVALH1]
“Foram os meus pais que me deram essa ideia”
- Uso de formas aparentemente finitas em contextos que requerem formas do infinitivo e vice-versa:
PA: *pode compra livro depois mais tarde* [CELESH3]
“pode comprar o livro depois, mais tarde”
- Variação no emprego das preposições, que pode passar pela omissão destas, pela aplicação de formas redundantes ou pelo uso do item “ni”, em substituição de “em”:
PA: *o quê você faz de manhã até na hora que ocê vai dormir de noite?* [OSVALH1]
“o que é que você faz desde manhã até à hora em que você vai dormir à noite”
- Redução de concordância verbal, que atinge todas as pessoas do verbo:

PA: *nós compra coesa bucado barato* [FRANCM2]

“nós comprávamos as coisas um bocado baratas”

Tendo em conta o tipo de fenômenos linguísticos citados, pode-se afirmar que o PA reestruturado é proveniente de um processo de transmissão linguística irregular, pois

A sociohistória da comunidade de Almojarife, por seu lado, concede evidências que permitem relacionar, empiricamente, tais fenômenos quer com o processo massivo, defectivo e não normatizado de aquisição do português, marcado por acentuada alteração nos seus parâmetros sintáticos e levado a cabo por escravos africanos geográfica e culturalmente transplantados para o entreposto de São Tomé [...] Inicialmente, devido à situação de contacto restrito relativamente ao PE, o PA não teve substancial acesso aos modelos deste, o que determinou que o seu processo de reestruturação ocorresse sem referências normatizadoras. Com campo aberto para a fixação de estruturas desviantes, o PA incorporou variação de diferentes tipos, que foi transmitida de geração em geração. (FIGUEIREDO, 2010, p. 98)

Ainda é possível perceber que os fenômenos de variação presentes no PA não são encontrados no PE, mas sim em outras variedades do português, como o PB (cf. LUCCHESI, BAXTER e RIBEIRO, 2009), que também emergiu de uma situação de contato linguístico. Portanto, através da análise variacionista e comparativa da CV-P4 em São Tomé, será possível observar até que ponto os fatores que orientam essa variação no PST e no PB popular se assemelham.

2.1.2 Origem dos dados; estrutura do *corpus*

A recolha das gravações que compõem o *corpus* deste trabalho foi feita sob a perspectiva sociolinguística (LABOV, 1972), foram coletadas 18 entrevistas entre fevereiro e maio de 1998 no projeto “Semi-creolization: testing the hypothesis against data from Portuguese – derived languages of São Tomé”, financiado pela Australian Reserch Council e dirigido por Alan Norman Baxter com colaboração de Dante Lucchesi.

O responsável pelas gravações das entrevistas, bem como pelas transcrições das mesmas, em Almojarife, foi Dante Lucchesi, contando com a intermediação com a comunidade feita pelo especialista em malária, Dr. Derek Charlwood, da Universidade de Londres, que tinha desenvolvido parte do seu próprio projeto nessa mesma comunidade.

Para a devida interpretação das produções de fala dos informantes nas transcrições dispôs-se da ajuda de Francisco Paulino. Posteriormente, na Universidade de Macau, todas as

transcrições foram revistas e digitalizadas por Carlos Figueiredo, que nasceu em Angola e tem familiaridade com variedades de português dessa região.

Para o presente trabalho, foram utilizadas as 18 entrevistas do corpus, seguindo os seguintes passos:

- Leitura das transcrições para familiarização com a variedade linguística em questão e com o contexto cultural;
- Levantamento dos dados;
- Codificação dos dados;
- Levantamento das hipóteses;
- Rodagem e análise dos dados.

O corpus da comunidade de fala de Almojarife é dividido por sexo e idade, e por nível de escolarização e estada fora da comunidade, uma vez que esses fatores sociais têm se mostrado relevantes para os estudos sociolinguísticos do PB. Quanto à profissão, os informantes exercem as mais variadas, apresentando também uma diversidade quanto ao nível de escolarização, que vai desde a ausência até a graus incompletos ou completos. Vale salientar que todos os informantes são bilíngues em santome e português reestruturado. O quadro 5 mostra os dados dos informantes.

Quadro 5. Informantes de Almojarife: faixas etárias e seus aspectos sociais. Obs. Os nomes dos informantes não são fornecidos para privacidade da identidade dos mesmos (FIGUEIREDO, 2010, p. 309).

Nr.	Código do informante	Faixa etária	Sexo	Escolaridade	Estada fora da comunidade	Duração da entrevista
01	[OSVALH1]	FE1(20-40)	Masc.	4ª. Classe	Não	01:00:28
02	[ZECAH1]	FE1 (20-40)	Masc.	4ª. Classe	Não	01:10:00
03	[CASTEH1]	FE1 (20-40)	Masc.	8º. Ano	Não	00:59:14
04	[ANTOM1]	FE1 (20-40)	Fem.	8º. Ano	Não	01:04:45
05	[DULCEM1]	FE1 (20-40)	Fem.	3ª. Classe	Não	00:56:20
06	[CARMOM1]	FE1 (20-40)	Fem.	8º. Ano	Não	01:06:19
07	[ABILH2]	FE2 (41-60)	Masc.	4ª. Classe	Sim (Gabão e Libreville)	01:08:59

08	[LUIH2]	FE2 (41-60)	Masc.	Não	Sim (Príncipe)	01:25:57
09	[MAURIH2]	FE2 (41-60)	Masc.	3ª. Classe	Sim (Libreville)	01:03:28
10	[CLOTIM2]	FE2 (41-60)	Fem.	4ª. Classe	Não	01:02:36
11	[FRANCM2]	FE2 (41-60)	Fem.	Não	Não	00:56:28
12	[MANEVEM2]	FE2 (41-60)	Fem.	Não	Sim (Gabão)	01:01:10
13	[CELESH3]	FE3 (+60)	Masc.	3ª. Classe	Não	01:04:44
14	[MANOH3]	FE3 (+60)	Masc.	Não	Não	01:02:55
15	[SALVH3]	FE3 (+60)	Masc.	Não	Não	01:10:53
16	[PRINCEM3]	FE3 (+60)	Fem.	Não	Não	01:15:05
17	[MMDEUSM3]	FE3 (+60)	Fem.	Não	Sim (Angola)	01:11:51
18	[MINISM3]	FE3 (+60)	Fem.	1ª. Classe	Não	01:08:16

2.2 ENQUADRAMENTO TEÓRICO

A análise dos dados nesta pesquisa segue os pressupostos da Sociolinguística Quantitativa, proposta por Labov (2008 [1972]; 1982; 1994), a qual aborda a língua como heterogênea e condicionada por fatores linguísticos e sociais em uma determinada comunidade de fala¹⁰. A pesquisa sociolinguística é direcionada para as variações sistemáticas concernentes ao seu objeto de estudo concebidas como uma heterogeneidade estruturada.

As formas em variação recebem o nome de "variantes linguísticas". Labov (2008 [1972], p. 221) afirma que variantes linguísticas são cada uma das maneiras de se dizer a mesma coisa em um mesmo contexto, isto é, com o mesmo valor de verdade.

Essas variáveis são divididas em dependentes e independentes. A variável dependente é o fenômeno a ser estudado, como a aplicação da regra de CV-P4, por exemplo. As variantes seriam as formas que estão em competição: a presença ou a ausência da regra de CV.

O uso de uma ou outra variante é motivado por fatores linguísticos ou extralinguísticos, conforme Mollica (2003, p. 10) "toda variação é motivada, isto é, controlada por fatores de maneira tal que a heterogeneidade se delinea sistemática e previsível". São os fatores linguísticos ou extralinguísticos que estabelecem as *variáveis explanatórias ou independentes*.

Por meio da análise das variáveis sociais, procura-se explicar qual o quadro de variação presente na comunidade de fala nos termos da dicotomia entre *variação estável* e *mudança em progresso*. Por *variação estável* se entende quando uma variação entre diferentes formas linguísticas pode se manter por um período de tempo porque ainda não é verificada uma tendência no uso de uma variante linguística sobre a outra. *Mudança em progresso* se trata de quando o processo de variação estável começa a se findar e uma das variantes concorrentes tem seu uso favorecido na comunidade de fala ao passo em que há uma diminuição na frequência da outra variante.

Através dessa dicotomia, pode-se observar de maneira clara os problemas da mudança linguística: o problema da transição, analisando como uma determinada variante se propaga na comunidade de fala; o problema da avaliação, o qual pode ser visto mediante o comportamento da comunidade de fala diante da variável observada; o problema do encaixamento, no qual se pode verificar informações de como uma estrutura linguística se

¹⁰ Comunidade de fala para esse modelo teórico-metodológico é entendida como um grupo de pessoas que compartilham traços linguísticos que distinguem seu grupo de outros. (cf. LABOV, 2008 [1972]).

firma dentro de uma comunidade de fala em duas vias: (i) encaixamento linguístico (estrutural) e (ii) encaixamento social; chegando ao problema da implementação, o que seria a etapa em que o processo de mudança se concretiza. (WEINREICH, LABOV E HERZOG, 2006 [1975], p. 124).

Dessa forma, é necessário observar se o quadro de variação indica o uso de duas formas linguísticas concorrentes ou se esse quadro já se configura como uma mudança linguística. Para verificar esse pressuposto, deve-se considerar que a variação observada no plano sincrônico em um determinado ponto da estrutura da gramática de uma comunidade de fala pode refletir um processo de mudança em curso na língua, no plano diacrônico. Assim, é no estudo em *tempo real*, que ocorre o desenvolvimento diacrônico da língua, no chamado estudo em *tempo aparente*. A proposição do tempo aparente é a de que é possível captar mudanças através da análise do comportamento linguístico de diferentes gerações. (LABOV, 1994, p. 83).

Além dos pressupostos teórico-metodológicos da Sociolinguística Quantitativa Laboviana, na discussão dos motivos que fundamentariam a variação da CV-P4 no PA, esta pesquisa se servirá da perspectiva teórica sobre o contato entre línguas e aquisição de L2 e L1. Pelo fato de haver bilinguismo em Almocharife, é relevante considerar os processos de aquisição de L1 e L2 para a formação do PA.

Sendo assim, a análise dos resultados deste trabalho é orientada também pela noção da transmissão linguística irregular (BAXTER, 1995; BAXTER E LUCCHESI, 1997; LUCCHESI, 2003; 2009). Essa teoria sociolinguística leva em conta as situações de contato que envolvem mudança de língua por meio da aquisição de L2, com etapas de bilinguismo, e a aquisição de L1 com *input* de dados de L2, aspectos relacionados à GU (cf. seção 1.3 deste trabalho).

2.2.1 Aspectos metodológicos

Para a quantificação dos dados foi usado o programa computacional GoldVarb (2001). O GoldVarb funciona a partir da formulação de regras variáveis, “um tipo de análise multivariada” que tem como objetivo “separar, quantificar e testar a significância dos efeitos de fatores contextuais em uma variável linguística.” (GUY; ZILLES, 2007, p. 33-34).

O GoldVarb compreende o pacote de programas VARBRUL, que é “um conjunto de programas computacionais de análise multivariada, especificamente estruturado para acomodar dados de variação sociolingüística” (GUY e ZILLES, 2007, p. 105). Este, partir de

uma variável dependente dotada de dois ou mais valores representativos do fenômeno a ser analisado (no caso deste trabalho a CV-P4) e as variáveis independentes (linguísticas e extralinguísticas), combina relações entre variável dependente e variáveis independentes para determinar como os fatores das independentes motivam o fenômeno representado pela dependente.

Os resultados são obtidos em pesos relativos que quantificam a interferência de cada variável independente na aplicação do fenômeno em análise, quais os que contribuem ou não para a escolha da variante (para o favorecimento ou desfavorecimento).

Na observação do corpus da comunidade de Almojarife, foi feito o levantamento dos dados. Esta etapa consiste em procurar e destacar nas transcrições feitas o objeto de estudo em questão, as ocorrências de CV-P4. Foram observadas as frases no contexto em que as variantes apareceram para verificar quais os fatores condicionadores da variação. Cada fator é constituído por subfatores e para cada subfator deve-se atribuir um valor representado por letra ou número (processo chamado de codificação), pois isso facilita a quantificação dos dados.

Os dados coletados para a análise da CV-P4 foram codificados levando em conta as matrizes analíticas adotadas em trabalhos anteriores sobre esse tema, incorporadas na chave de codificação elaborada pelo grupo Vertentes do ILUFBA (Instituto de Letras da Universidade federal da Bahia). Portanto, para o tratamento dos dados, foram contempladas as variáveis dependentes listadas na chave de codificação a seguir, assim como as hipóteses sobre o condicionamento dessas variáveis (subseção 3.2), representadas pelas variáveis independentes (1) a (8), da mesma lista.

VARIÁVEL DEPENDENTE

1. Ausência x presença de concordância sujeito-verbo

(f) presença (-mos, -mo, -emo)

(z) ausência

VARIÁVEIS LINGUÍSTICAS EXPLANATÓRIAS

1. Realização e posição do pronome sujeito

(R) realizado imediatamente antes do verbo (ex.: *Nós ficô chateado.*)

(D) realizado antes do verbo, mas separado por algum constituinte (ex.: *Nós ainda num fez não*)

(P) realizado e posposto (ex.: *Só ficô a gente na festa / Só ficamos nós na festa.*)

(E) pronome sujeito não realizado (ex.: Nós não saimo. *Ficamo em casa.* / *A gente não saiu. Ficô em casa.*)

(Q) retomado por pronome relativo (*A gente, que trabalha aqui, folga na sexta.*)

2. Nível de referencialidade (significado)

(e) eu ([+ específico])

(V) eu + você(s) ([+ específico])

(o) eu + ele(s) – sem você (s) ([+ específico])

(t) eu + você(s) + ele(s) ([+específico])

(G) grupo de contornos pouco definidos (- específico])

(I) genérico ([- específico])

3. Paralelismo formal

(a) forma isolada ou primeira referência

(b) precedido por *nós* + verbo com desinência – *mos/-mo*, na oração anterior

(c) precedido por *nós* + verbo com desinência ϕ , na oração anterior

(d) precedido por verbo com desinência – *mos/-mo*, sem sujeito realizado, na oração anterior

(e) precedido por *a gente* + verbo com desinência – *mos/-mo*, na oração anterior

(f) precedido por *a gente* + verbo com desinência ϕ , na oração anterior

(h) precedido por verbo com desinência ϕ , sem sujeito realizado, na oração anterior

4. Saliência fônica (NARO *et all*, 1999, p. 203)

(1) Nível 1: tanto no singular quanto no plural do verbo, não há mudança da posição da sílaba tônica.

(ex.: *falava/falávamos; fosse/fôssemos; ia/íamos*)

(2) Nível 2: há mudança na posição da sílaba tônica quando o verbo muda do singular para o plural.

(ex.: *fala/falamos; trouxe/trouxemos; assiste/assistimos; fique/fiquemos*)

(3) Nível 3: a sílaba tônica é a mesma no singular ou no plural, porém há o acréscimo de segmentos na forma plural.

(ex.: *está/estamos; tem/temos; fazer/fazemos; faz/fazemos; será/seremos; dá/damos; vê/vemos*)

(4) Nível 4: a posição da sílaba tônica no verbo é a mesma no singular e no plural, porém na primeira forma ocorre uma semivogal que não ocorre na segunda.

(ex.: *comeu/comemos; partiu/partimos; vai/vamos; foi/fomos*)

- (5) Nível 5: a posição da sílaba tônica é a mesma, mas ocorre mudança no radical.
(ex.: *veio/viemos; é/somos; fez/fizemos*)

VARIÁVEIS SOCIAIS

5. Faixa etária

- (1) faixa I (20 - 40 anos)
(2) faixa II (21 - 60 anos)
(3) faixa III (acima de 61 anos)

6. Sexo

- (H) masculino
(M) feminino

7. Escolarização

- (\$) semi-analfabeto
(@) analfabeto

8. Estada fora da comunidade

- (*) viveu pelo menos seis meses fora da comunidade
(#) o informante sempre viveu na comunidade

Foi feita a codificação de todas as ocorrências de CV-P4. Seguem-se exemplos de ocorrências codificadas:

- (7) (zRIc22M@# *nós procura folha de banana* [ANTOM1]
(8) (zRIc13H@# *nós podia tê luz aqui já* [MANOH3])

2.3 AS HIPÓTESES DO TRABALHO

Na análise de P4 nesse trabalho a variável dependente foi assim determinada: (f) presença da marca de concordância, exemplificada em (9), (10) e (11) e (z) ausência da marca de concordância, exemplo (12).

- i) (9) *nós trabalhamos lá* [MAURIH2]
(10) *nós stamo fala* [MANEVEM2]
(11) *nós puxemos laço* [ZECAH1]

ii) (12) *nós ta nós assim só* [DULCEM1]

Com base nos estudos prévios sobre a concordância P4 no português brasileiro, foram levantadas 9 hipóteses sobre o condicionamento da variável dependente no português de STP, que serão avaliadas por 8 variáveis independentes, sendo 4 linguísticas e 4 extralinguísticas. Passamos a descrever a motivação de cada variável.

A hipótese geral sobre a variável dependente, a CV-P4, na fala dos almoxarifanos é que sejam encontrados mais dados com ausência do que com presença de concordância. Essa suposição é baseada no fato de seus informantes possuírem baixo grau de escolaridade e pertencerem a uma comunidade com características rurais, além do fato de serem falantes também de uma língua em que não há flexão de concordância.

A variável independente (1) **realização e posição do sujeito** avalia o comportamento do sujeito quanto à sua posição e realização influencia na aplicação da regra de CV. As hipóteses levantadas são duas: (i) quanto menor a distância entre o sujeito e o verbo, mais a CV tenda a ser realizada; e, (ii) a não realização fonética do sujeito (sujeito nulo) favoreça a realização da concordância (LEMLE E NARO, 1977; NARO E SCHERRE, 2007; LUCCHESI, 2009).

(13) *nós stamos nesse tempo aqui ão vala pena* [CLOTIM2]

(14) *Depois vivíamos lá* [ANTOM1]

A variável (2) **nível de referencialidade**, diz respeito à referência do falante quando utiliza “nós”. Neste caso, a hipótese de trabalho é que quanto mais o sujeito apresente o traço específico (eu ([+ específico])), maior a probabilidade de aplicação da regra, enquanto se esse referente for mais genérico ([- específico]), é provável que a concordância tenda a não se realizar.

(15) *eu mia mia mãe só somos dois* [CLOTIM2]

A variável (3) **paralelismo formal** ou **paralelismo discursivo** diz respeito à questão de que formas parecidas tendem a permanecer juntas: marcas levam a marcas e zeros levam a zeros (NARO, 1993). Com base nesse princípio, as hipóteses para essa variável são duas: (i) que a presença de uma forma verbal marcada na oração anterior favoreça a aplicação da regra de CV; (ii) devido a fatores funcionais, o morfema de pessoa usado na primeira referência

também favoreça a CV, visto que na primeira referência do sujeito (P4), as marcas de plural tornam-se mais necessárias.

(16) *nós eramos 5 numa canoa, fomos dentro de canoa* [DULCEM1]

(17) *nós stamos nesse tempo aqui ñ vala pena* [CLOTIM2]

A variável (4) **saliência fônica** segue mesma a hipótese de estudos já discutidos anteriormente: quanto mais saliente for a diferença entre singular e plural, mais provável seja a aplicação da regra de concordância. Utilizando a escala proposta de Naro *et all* (1999, p.203), são cinco níveis para a saliência fônica:

(1) a posição da sílaba tônica é a mesma tanto na forma singular quanto no plural;

(2) a sílaba tônica muda de posição quando o verbo passa do singular para o plural, ocorrendo nasalidade no plural;

(3) a sílaba tônica é a mesma quando verbo está no singular ou no plural;

(4) a posição da tonicidade é a mesma quando verbo está o singular ou no plural, mas no primeiro ocorre uma semivogal que não ocorre no segundo;

(5) a posição da tonicidade é a mesma no plural ou no singular, porém ocorre uma mudança no radical do verbo.

Os níveis mais propícios à aplicação da CV são os níveis 4 e 5, exemplos (18) e (19), porque envolvem mudanças diversas na forma plural: mudanças vocálicas na desinência, mudanças na raiz, e até mudanças completas.

(18) *Somos 2 nessa semana* [CASTEH1]

(19) *nós já fomos a equipa uma vez* [CASTEH1]

As variáveis sociais estão interligadas ao comportamento linguístico da variação e da mudança, visto que entram nesse contexto a noção de estigma e de prestígio (LUCCHESI E ARAÚJO, 2004, s/p). A variável (5) **faixa etária** permite detectar se na comunidade ocorre mudança em curso ou mudança em progresso, porém esta sempre se associa a outros fatores, como a escolarização. Levando em consideração as diferentes gerações de informantes, distribuídas em três faixas etárias.

A variável (6) **sexo do falante** nos estudos brasileiros apresentam duas tendências: (i) que as mulheres impulsionam a mudança utilizando a variável de prestígio quando se trata de comunidades urbanas (CHAMBERS E TRUDILL, 1980, p. 97-8); (ii) na fala de comunidades

rurais há uma tendência de que os homens liderem a mudança para a variável de prestígio (LUCCHESI, 2009, p. 369). No caso da comunidade de Almojarife, busca-se saber qual é o efeito do sexo do falante sobre a variável dependente. Portanto, a hipótese é: o sexo do falante vai condicionar o uso da variável dependente.

Em alguns estudos prévios, sobre o PB, a variável (7) **escolaridade**, tem revelado um determinado tipo de condicionamento, com a tendência de que os falantes com maior nível de escolarização utilizam mais a norma padrão, enquanto a classe intermediária e classe mais baixa com menor grau de escolaridade conservam os traços linguísticos mais populares (NARO, 1981). A respeito da comunidade de Almojarife, propõe-se essa variável conforme a hipótese de que o nível de escolarização condicione a variação da CV-P4.

A variável (8) **estada fora da comunidade** tem revelado uma tendência para a aquisição da norma padrão através da influência dos centros urbanos a partir de indivíduos que viveram fora das comunidades rurais por um determinado período de tempo (LUCCHESI, 2009, p. 369). Em relação à comunidade de Almojarife, acredita-se que não somente a urbanização, mas o contato no exército colonial e do trabalho nas grandes roças condicionem a aplicação da regra de CV-P4.

Foram codificadas e contabilizadas um total de 542 ocorrências da CV-P4 para processamento pelo programa GoldVarb. A análise e a interpretação dos dados são apresentadas no capítulo 3 deste trabalho. Vale salientar que para cotejar os resultados desta pesquisa com os de outras realizadas sobre o PB, foram escolhidos dois trabalhos sobre comunidades de fala cujas características se aproximam deste estudo para o PST, mais precisamente, para o PA, e tratam da CV-P4, a saber: Almeida (2006) e Lucchesi (2009).

Ambos tratam da CV-P4. O primeiro estudo apresenta dados referentes a uma comunidade de fala de afro-brasileiros situada em Restinga Seca, no Rio Grande do Sul, a comunidade de São Miguel dos Pretos. O segundo estuda comunidades afro-brasileiras do interior da Bahia (Rio de Contas, Sapé, Cinzento e Helvécia). A escolha desse trabalhos foi motivada pelo fato de tratarem de comunidades povoadas por escravos e seus descendentes, falantes de português com forte influência de línguas africanas, além do fato de abordarem aspectos linguísticos semelhantes aos que foram propostos nesta dissertação sobre o PA.

2.4 SÍNTESE

Neste capítulo foi apresentado o enquadramento metodológico do presente trabalho. Apresentou-se a comunidade de falares na seção 2.1, seus aspectos socio-históricos, bem como suas características sociolinguísticas. Também foram mostrados os dados dos informantes e como estes foram caracterizados (subseção 2.1.1).

A subseção 2.1.2 tratou da origem dos dados e da estrutura do *corpus*, mostrando como se deu a recolha dos dados, seguida da seção 3 que descreveu os procedimentos metodológicos deste trabalho, salientando a utilização da Sociolinguística Quantitativa (LABOV, 1972) como modelo de análise. Pois, ela permite determinar o modo como a CV-P4 se encaixa na gramática e na sociedade, isto é, permite entender quais as motivações que determinam como esse fenômeno é percebido socialmente.

A análise dos dados foi feita através de métodos matemáticos quantificados pelo VARBRUL que pertence ao recurso do pacote computacional GoldVarb (2001), que permitiu identificar de forma sistemática as variáveis relevantes para o fenômeno da CV-P4 na fala dos almoxarifanos, subseção 2.2.1.

Nessa mesma subseção foi apresentada a chave de codificação utilizada para a rodagem dos dados no GoldVarb (2001) e a variável dependente e as independentes que correspondem às hipóteses apresentadas em 2.3. A apresentação e discussão dos resultados da análise é mostrada no capítulo 3.

CAPÍTULO 3 - ANÁLISE DOS DADOS

Neste capítulo são apresentados os resultados obtidos através da análise quantitativa dos dados da CV-P4 no corpus da comunidade de fala de Almojarife em São Tomé e Príncipe. O capítulo está dividido em três seções. Na seção 3.1, são apresentados os dados gerais desta pesquisa, como se deu a análise quantitativa, bem como as variáveis inicialmente levantadas e quais as selecionadas e não-selecionadas pelo VARBRUL.

A seguir, na seção 3.2, são abordados os aspectos do encaixamento estrutural do fenômeno, em relação às variáveis linguísticas selecionadas como estatisticamente relevantes para a aplicação da CV-P4. Finalmente, na seção 3.3, apresenta-se os resultados das variáveis sociais e discute-se o encaixamento social da CV-P4. Ao longo de toda a discussão dos resultados é feita uma comparação com outros estudos sobre a variável P4.

3.1 RESULTADOS GERAIS

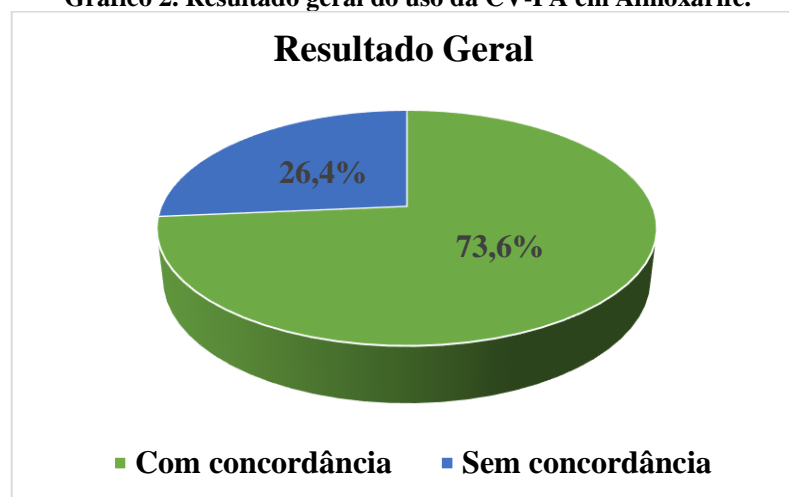
Após o levantamento e a codificação dos dados, seguindo o aparato metodológico descrito no capítulo 2, os dados codificados foram quantificados e preparados para a análise de multivariáveis.

A avaliação inicial da distribuição revelou um total de 542 ocorrências da variável dependente, sendo que houve 399 ocorrências com concordância, equivalendo a 73,6% dos dados, e 143 sem concordância, equivalendo a 26,4%, como se pode ver na Tabela 4 e no Gráfico 2.

Tabela 4. Frequência de aplicação da regra da CV-P4 na fala da comunidade de Almojarife em São Tomé.

Variante	Ocorrências/Total	Frequência
com concordância	399/542	73,6%
sem concordância	143/542	26,4%

Gráfico 2. Resultado geral do uso da CV-PA em Almojarife.



Considerando a frequência global de 73,6% de concordância, algumas questões já podem ser destacadas. Primeiro, acreditava-se que os resultados da presença de concordância na fala da comunidade de Almojarife seria menor do que a ausência em virtude de seus informantes possuírem baixo grau de escolaridade e pertencerem a uma comunidade com

características rurais, além do fato de serem falantes também de uma língua em que não há concordância. No entanto, os dados mostram que a concordância faz parte da gramática dessa comunidade.

Acredita-se que a comunidade de Almoxarife está adquirindo a forma padrão referente à P4 devido à ação das variáveis sociais, como a faixa etária e a escolaridade. Os falantes mais novos, por possuírem mais escolarização, têm ou tiveram mais contato com a variedade padrão da língua e tiveram a oportunidade de adquiri-la.

Por outro lado, os mais velhos são os que tiveram menos chances de frequentar a escola, logo, a progressão no uso do padrão é inversamente proporcional à idade, o que leva a inferir que a aquisição da marca de CV está ocorrendo de geração em geração (cf. subseções 3.3.1 e 3.3.3). Além disso, alguns informantes trabalham na venda e compra de produtos, assim, a atividade comercial e o conseqüente contato com pessoas de outras comunidades, tanto da zona rural quanto da zona urbana, pode justificar a alta frequência de CV.

Segundo, nota-se uma disparidade desses resultados com os obtidos em pesquisas em comunidades brasileiras cujos informantes possuíam pouca ou nenhuma escolaridade. A frequência de 73,6% de aplicação da regra de CV - P4 registrado aqui difere como, por exemplo, do estudo de Lucchesi (2009, p. 364) que apresentou apenas 18% de frequência. Essa diferença surpreende quando se lembra que a aquisição do português na comunidade de Almoxarife é relativamente recente (século XX, essencialmente).

Por outro lado, o perfil registrado aqui assemelha-se ao registrado no estudo de Almeida (2006, p. 89-90) com falantes com pouca ou nenhuma escolarização que realizam a concordância de P4 com uma frequência de 73% de aplicação da regra e 27% de não aplicação. Almeida justifica a alta frequência de realização de concordância devido ao fato de os moradores estabelecerem laços de trabalho com antigos fazendeiros da região e por manterem diferentes relações com pessoas de outras comunidades, tanto na zona rural quanto na zona urbana, devido à necessidade de emprego, saúde e educação.

A quantificação preliminar da variável dependente revelou a presença de KNOCKOUTS em dois grupos de fatores. Assim, por um lado, no grupo (3), **realização e posição do pronome sujeito**, os fatores <P> *sujeito realizado e posposto* e <Q> *retornado por pronome relativo* não manifestavam variação e apresentavam, respetivamente 2 e 3 dados. Por outro lado, no grupo (5), **paralelismo formal**, relevaram-se KNOCKOUTS representados por poucos dados nos fatores <f> *precedido por a gente + verbo com desinência ϕ , na oração anterior*, <h> *precedido por verbo com desinência ϕ , sem sujeito realizado, na oração anterior* e <n> *precedido por a gente + verbo com desinência – mos/-mo, na oração anterior*.

Após eliminar os KNOCKOUTS, optou-se por verificar a ortogonalidade da base de dados, examinando o cruzamento de cada par de variáveis independentes por meio de CROSSTAB, procedimento essencial prévio à análise de multivariáveis (GUY & ZILLES, 2007, p. 34; LABOV 2001, P. 85; TAGLIAMONTE 2012, p. 121).

O CROSSTAB revelou uma série de lacunas na distribuição dos dados, lacunas não aparentes em uma simples rodada de CELL. Dessa maneira, por um lado, descobriu-se que na variável independente (2), **nível de referencialidade**, apenas dois fatores, fatores <o>, eu + ele(s) – sem você(s) [+ específico], e <i>, genérico [- específico] apresentam ortogonalidade com os demais grupos, o que torna o grupo (4) inviável, pois os fatores <o> e <i> são inadequados para avaliar o efeito dessa variável. De maneira semelhante, no grupo (3), apenas dois fatores, <a>, forma isolada ou primeira referência, e <d>, precedido por verbo com desinência – *mos/-mo*, sem sujeito realizado, na oração anterior, apresentam ortogonalidade com os demais grupos de fatores.

Essa falta de ortogonalidade também tornou inviável o grupo (3), pois os fatores <a> e <d> não servem para avaliar essa variável de maneira adequada desde que não há termos de comparação. Se não há um fator para avaliar o efeito da ausência de flexão no verbo anterior, é impossível avaliar se a ausência de flexão no verbo anterior desfavorece a flexão no verbo em análise.

Por outro lado, foram reveladas interações entre o grupo (5), faixa etária, e os grupos (7), escolaridade, e (8), estada fora da comunidade. Efetivamente, o grupo (9), **escolaridade**, apenas apresenta ortogonalidade com o grupo (5), **faixa etária**, nas faixas 2 e 3, sendo que todos os informantes da faixa 1 são escolarizados. Da mesma maneira, a variável (8), estada fora da comunidade, não se aplica à faixa etária 1, pois nenhum informante nessa faixa viveu fora da comunidade.

Diante dessa realidade, por serem inoperacionais, foram eliminados da análise os grupos (2), **nível de referencialidade**, e (3), **paralelismo formal**. Depois, foram realizadas três rodadas de VARBRUL:

(i) Na primeira rodada (R1), optou-se por excluir da análise os grupos (7), **escolarização**, e (8), **estada fora da comunidade**, devido à falta de ortogonalidade com o grupo (8), **faixa etária**.

(ii) Em uma segunda rodada (R2), os grupos (7) e (8) foram restaurados, ambos sem os dados da faixa 1 do grupo (6), sexo; e o grupo (6) foi retirado da rodada.

(iii) Finalmente, fez-se uma terceira rodada (R3). O grupo (8), **estada fora da comunidade**, foi retirado do por ter manifestado interferência com o grupo (8), **faixa etária**;

e os grupos (6) e (7), **escolarização**, foram incluídos, ambos sem os dados da faixa 1 do grupo (6).

O Quadro 6 apresenta um resumo das três rodadas, mostrando as variáveis selecionadas e a respectiva ordem de seleção das mesmas:

Quadro 6. Variáveis independentes processadas pelo VARBRUL.

Variáveis independentes ↓	RODADA 1		RODADA 2		RODADA 3	
	Variável selecionada	Ordem de seleção	Variável selecionada	Ordem de seleção	Variável selecionada	Ordem de seleção
Realização e posição do sujeito	+	3	+	2	+	2
Saliência fônica	+	4	+	4	+	3
Faixa etária	+	1	Omitida	Omitida	-	-
Sexo	+	2	-	-	-	-
Escolarização	Omitida	Omitida	+	1	+	1
Estada fora da comunidade	Omitida	Omitida	+	3	Omitida	Omitida

Nas próximas seções, 3.2 e 3.3, apresentam-se os resultados dessas três análises.

3.2 O ENCAIXAMENTO ESTRUTURAL

Na análise VARBRUL do encaixamento estrutural da variação da CV-P4 na gramática do português da comunidade de Almojarife, entre as variáveis linguísticas, foram estatisticamente relevantes as seguintes: **realização e posição do sujeito** e **saliência fônica**, selecionadas nessa mesma ordem. A variável **paralelismo formal**, que não foi avaliada pelo VARBRUL, pelos motivos explicados na seção anterior, será comentada pelo potencial interesse das frequências gerais que ela revela.

Nas seções seguintes são comentados os resultados dessas variáveis, considerando as hipóteses levantadas.

3.2.1 Variável *realização e posição do sujeito*

Com esta variável buscou-se ver se a forma e a posição do sujeito na oração, em relação ao verbo, interferem no fenômeno da concordância. Para esta análise, o grupo foi estruturado levando em conta os seguintes fatores: sujeito realizado imediatamente antes do verbo; sujeito realizado antes do verbo, mas separado por algum constituinte; sujeito realizado e posposto ao verbo; sujeito não realizado; sujeito retomado por pronome relativo.

Contudo, os fatores *sujeito realizado e posposto ao verbo* e *sujeito retomado por pronome relativo* foram retirados da análise final por apresentarem apenas dois e três dados, respectivamente. Os dados seguintes exemplificam os fatores envolvidos neste grupo:

(20) *veio nós neu ele* [MINISM3]

(21) *nós é que sforçamos* [ABILH2]

A Tabela 5 apresenta os resultados da análise efetuada pela primeira rodada:

Tabela 5. Efeito da *realização e posição do sujeito* sobre a variável dependente.

Sujeito	Nº de oco./Total	Freq.	RODADA 1 P.R.
Não realizado	150/180	83,3%	0.657
Realizado antes do verbo, não adjacente	42/76	55,3%	0.355
Realizado antes e adjacente ao verbo.	205/281	73%	0.437
TOTAL	397/537	73,6%	

Nível de significância: = 0.002

De acordo com os resultados apresentados na Tabela 5, a não realização do sujeito favorece à aplicação da CV, com PR .657 e uma frequência de 83, 3%. Sendo assim foi confirmada a hipótese de que a CV seria mais realizada quando não houvesse a realização fonética do sujeito. Isso demonstra que esses falantes têm (alguma) sensibilidade relativa à relação sujeito-verbo, pois eles percebem que a desinência verbal é um indicador de sujeito. Relativamente ao efeito do apagamento do sujeito no português afro-brasileiro, Silva (2009) afirma que o *princípio da coesão estrutural*, influencia a concordância verbal quando o sujeito não é realizado, uma vez que

Sumariamente, esse princípio prevê que, numa situação em que há concorrência entre duas gramáticas na competência linguística do falante, o predomínio de uma gramática sobre a outra se dá em porções determinadas da estrutura da frase. No caso das comunidades rurais afro-brasileiras isoladas, observa-se a concorrência entre uma gramática mais crioulezante (sem regras de concordância), ligada às origens históricas dessas comunidades, e uma gramática mais padrão (com regras de concordância), que estaria penetrando na comunidade em função das influências externas. Essa concorrência se manifesta na produção verbal de cada falante da comunidade, sendo restringida pelo princípio da coesão estrutural, de modo que, por exemplo, se o falante inicia uma oração com a gramática padrão (o que implica a aplicação da regra de concordância nominal no SN sujeito), ele tende a aplicar a regra de concordância também na relação entre esse sujeito e o verbo, pois é o sujeito que especifica a concordância verbal. (SILVA, 2009, p. 352-353).

Dessa maneira, segundo o princípio da coesão estrutural, supõe-se que a regra gramatical padrão do português, que aceita sujeito não realizado, é evidenciada nos dados analisados. Logo, se o falante inicia uma frase com a gramática padrão com o sujeito não realizado na primeira pessoa do plural, a regra concordância verbal tende a ser aplicada no restante da sentença. Prevalece também uma lógica funcional (LUCCHESI, 2009, p. 364-365), já que o morfema de número e pessoa supriria a ausência de informação com o sujeito nulo, como mostram os exemplos (22) e (23).

(22) *Depois vivíamos lá* [ANTOM1]

(23) *estamos assim sempre* [CASTEH1]

O *sujeito realizado imediatamente antes do verbo*, apesar de apresentar PR .437 e obter uma frequência de 73%, indicando um leve desfavorecimento da regra de concordância. Dessa forma, a hipótese de que o sujeito imediatamente anteposto ao verbo favorece a concordância verbal não se confirmou.

O *sujeito realizado, mas separado por algum interveniente* apresentou PR .355, medianamente desfavorável à aplicação da regra. O fato de o sujeito estar mais distante do verbo na estrutura frasal, deve fazer com que o falante tenha mais dificuldades para recuperar a referência do sintagma nominal para estabelecer a concordância com o verbo.

O fato de que os falantes, no *corpus* contemplado, serem bilíngues em português e em uma língua crioula (que não tem morfologia concordante) pode explicar essa diferença. Na gramática do crioulo não haveria diferença nenhuma entre um *sujeito presente, mas distante do verbo* ou um *sujeito adjacente ao verbo*, pois o verbo não tem flexão de concordância. Portanto, acredita-se que esses falantes no *corpus* estão orientados pela gramática do crioulo, nesse sentido.

No trabalho com os afro-brasileiros, de Lucchesi (2009, p. 365), a variável realização e posição do sujeito apresentou resultados semelhantes aos fatos observados aqui. O sujeito não realizado também foi o fator que mais favoreceu a CV - P4 com PR .88 e 38% de frequência.

O autor justifica esse resultado através da lógica funcional, pois o morfema de número e pessoa supriria a ausência de informação com o sujeito nulo. A presença do pronome sujeito imediatamente antes do verbo ou separado por algum constituinte desfavoreceu a aplicação da regra de concordância, com PR .38 e PR .37, e 13% e 11% de frequência, respectivamente.

Esses resultados contradizem as conclusões de Scherre e Naro (1993), com base em seus estudos da concordância verbal com a 3ª pessoa do plural, de que o “princípio do paralelismo formal”, segundo o qual “marcas levam a marcas e zeros levam a zeros”, suplanta o “princípio da economia” na “repetição de marcas formais”; do mesmo modo que desautorizam a sua generalização de uma “tendência em direção a marcas redundantes” no PB. (NARO; SCHERRE, 2007, p.177). (LUCCHESI, 2009, p. 367)

Em Almeida (2006, p. 104-106), foi encontrada uma frequência de 70% de aplicação da CV¹¹ quando o sujeito foi anteposto ao verbo, favorecendo a CV-P4. E o tipo de sujeito que favoreceu a regra foi o sujeito apagado com 83% de frequência, sendo explicado também pela lógica funcional de apresentar a informação que o sujeito não releva através do emprego da desinência número-pessoal.

¹¹ No estudo de Almeida (2006), a variável *posição do sujeito* não foi submetida ao VARBRUL para obtenção dos pesos relativos devido à falta de equilíbrio numérico entre os fatores, pois houve a necessidade de eliminar *knockouts* que surgiram nessa variável. O mesmo ocorreu com o *tipo de sujeito*.

Através dos resultados obtidos para o PA, acredita-se que prevaleça a afirmação de Lucchesi (2009, p. 367), a respeito da variedade brasileira de comunidades afro-brasileiras do interior da Bahia, também se aplica ao presente estudo.

3.2.2 Variável *saliência fônica*

Variável clássica nos estudos sobre concordância verbal, a *saliência fônica* foi significativa nos resultados deste trabalho. Partiu-se do pressuposto de que a oposição acentuada favorece a presença de marcas explícitas de plural nos verbos, contrário à oposição não-acentuada, com maiores probabilidades de marcas zero de concordância, questão confirmada nos resultados dos dados.

Assim, a hipótese inicial para esta variável foi de que quanto mais saliente fosse a oposição morfofonológica do verbo, maior a possibilidade de concordância e é justamente o que pode ser notado na Tabela 6.

Tabela 6. Efeito da variável *saliência fônica*, sobre a variável dependente.

Grau	Nº	%	RODADA 1 P.R.
1: <i>falava/falávamos</i>	38/68	55,9%	.263
2: <i>fala/falamos</i>	111/139	79,9%	.597
3: <i>está/estamos</i>	76/120	63,3%	.457
4: <i>vai/vamos</i>	121/149	81,2%	.531
5: <i>é/somos</i>	53/66	80,3%	.567
TOTAL	399/542	73,6%	

Nível de significância: 0.002

Utilizando a escala de diferenciação morfofonológica na oposição singular-plural proposta por Naro *et all* (1999, p. 203), verificou-se que os resultados obtidos não seguem completamente a gradação hierárquica. Os níveis 2, 4 e 5 são os que favorecem a concordância verbal, apresentando, respectivamente, PR de .597, .531, .567. e frequências de 79,9%, 81,2% e 80,3%

Os PR e as frequências dos níveis 4 e 5 já eram esperadas, e, portanto, corroboram a hipótese levantada. Tal fato é percebido pelo contraste de valores dos pesos relativos do nível

1 (PR .263) e do nível 5 (PR .567). Porém a frequência e o peso relativo do nível 2 resultou de maneira inesperada no PR favorecedor de .597, o que leva à necessidade de considerar a ação de outros fatores no índice de concordância.

(24) *nós **falamos** aqui falamo* [MAURIH2]

(25) *nós **vamos** dar uns passo* [CARMOM1]

(26) ***eramos** 3 a trabalhá aisso* [CELESH3]

O resultado do nível 2 de saliência fônica dos verbos pode ser justificado pelo fato de que o nível 2 da escala apresenta um contraste que envolve uma mudança na posição da tonicidade, isto é, a sílaba tônica muda de posição quando o verbo passa do singular para o plural, como em *fala/falamos* exemplo (24). Sendo os verbos no crioulo uniformes, o falante tende a marcar a diferenciação que é mais perceptível.

No nível 1, a posição da tonicidade é a mesma tanto na forma singular quanto no plural, não ocorrendo nasalidade. Em oposição a isso, nos níveis 3, 4 e 5, a sílaba tônica ocorre em posições diferentes quando o verbo passa do singular para o plural. Portanto, é mais natural para o falante perceber a diferenciação morfofonológica desses níveis, o que favorece à concordância, do que em 1.

Usando a mesma escala dos níveis de saliência (NARO *et all*, 1999, p. 203), Lucchesi (2009, p. 367) verificou que no português afro-brasileiro os resultados exibiram uma perfeita sintonia com a escala proposta, o que é notável, já que nem sempre essa variável exibe um comportamento escalar, apresentando por vezes um resultado contrastivo entre dois extremos.

Ele ainda destacou que a frequência de aplicação da regra aumenta bastante a partir do nível 3 de saliência, em que a oposição ocorre na sílaba tônica das duas formas em contraste; a partir daí, o PR salta de .25 para .71 e a frequência da concordância passa de 5% para a casa dos 20%.

Já Almeida (2006, p. 98) verificou que os resultados obtidos na fala da comunidade de São Miguel dos Pretos não seguem completamente a gradação hierárquica proposta por Naro *et all* (1999, p 203). A gradação se inicia com o PR de .00 no nível 1 e no nível 2 sobre para .70 e cai novamente no nível 3 com .29. Almeida explica esse resultado através da ação de outra variável linguística, a presença do pronome reto na função sujeito, segundo ela isso ocorre porque existe a hipótese de que esse tipo de sujeito favoreça a aplicação da CV.

De toda sorte, comparando esses dados com os achados para o PB pode-se retomar as colocações feitas por Guy (1981, p. 296) de que haveria no Brasil um processo de mudança

em direção à aquisição da morfologia flexional, a partir dos contextos mais salientes para os menos salientes, algo típico nos processos de descrioulização.

Nesse sentido, infere-se que no PA isso se difere, pois o que é saliente para o falante de Almojarife vai depender da fonologia e fonética do crioulo, sendo assim o que é saliente no PA não é necessariamente o que é considerado saliente na escala de Naro *et all* (1999, p 203).

3.2.3 Variável *paralelismo formal*

Como se explicou na seção 3.1, a variável paralelismo formal não foi processada pelo VARBRUL. No entanto, as frequências gerais parecem sugestivas e serão comentadas aqui. A Tabela 7 mostra a distribuição dos dados gerais dessa variável.

Tabela 7. Frequência geral da variável *paralelismo formal*.

Paralelismo formal do sujeito em relação ao verbo	Nº de oco./Total	Freq.
<a> forma isolada ou primeira referência	43/50	86%
<d> precedido por verbo com desinência – mos/-mo, sem sujeito realizado, na oração anterior	112/127	88,2%
 precedido por nós + verbo com desinência – mos/-mo, na oração anterior	236/237	99,6%
<c> precedido por nós + verbo com desinência ϕ , na oração anterior	6/110	5,5%
<f> precedido por a gente + verbo com desinência ϕ , na oração anterior	1/5	20%
<h> precedido por verbo com desinência ϕ , sem sujeito realizado, na oração anterior	0/12	0%
<n> ou <e> precedido por a gente + verbo com desinência – mos/-mo, na oração anterior	1/1	100%
TOTAL	399/542	73,6%

Diante do exposto na Tabela 7, parece que os fatores <a>, , <c> e <d> apresentam dados que pudessem ser submetidos a uma análise por VARBRUL. Porém, uma análise por meio de CROSSTAB, cruzando o grupo (3) com demais grupos, um por um, revela que os fatores e <c> não são ortogonais relativamente aos demais fatores. Ou seja, a variável (3) não tem representação ortogonal relativamente a diversos fatores nos outros grupos.

Mesmo assim, a distribuição na Tabela 7 poderia sugerir que o paralelismo formal tenha algum papel no condicionamento da variável dependente. Nesse sentido, nota-se que as frequências de flexão mais altas pertencem precisamente ao verbo precedido por um verbo flexionado (fatores <d> e , 99,6% e 88,2%, respectivamente.), exemplos (27) e (28), enquanto que a ausência de flexão no verbo anterior (fatores <c>, <f> e <h>) traz frequências de flexão mínimas. Ressalta-se também a frequência de 86% quando a forma verbal foi isolada ou de primeira referência, exemplo (29).

(27) *uma roda que nós fazíamos, criamos brasil, inglaterra, suíça, gabão, portugal, angola, não sei que.* [CARMOM1]

(28) *nós fizemos isso metemos a banana metemos* [ABILH2]

(29) *nós aqui conversamos* [MMDEUSM3]

Lucchesi (2009, p. 366) confirma o princípio do paralelismo discursivo, já que a presença de uma forma verbal marcada na oração anterior foi o fator que mais favoreceu a aplicação da regra de concordância verbal, com PR .72 e uma frequência de 48%. Destacando também que, por fatores pragmáticos e funcionais, o morfema de pessoa e número também foi bastante usado na primeira referência: PR. 62 e frequência de 19%, justificando que, na primeira menção do sujeito, as marcas de referência tornam-se mais necessárias.

3.3 ENCAIXAMENTO SOCIAL

Para analisar o encaixamento social na variação da CV-P4 na comunidade de fala de Almojarife foram propostas, inicialmente, as seguintes variáveis: (i) faixa etária (20-40 anos; 41-60; e mais de 60 anos), (ii) sexo, (iii) escolarização, (iv) estada fora da comunidade e (v) localização. Na primeira rodada dos dados, apenas as variáveis *faixa etária* e *sexo* foram selecionadas como significativas. No entanto, considerando a interação negativa entre *faixa etária* e *estada fora da comunidade*, para melhor compreensão do efeito das variáveis sociais, fez-se uma segunda rodada (R2) para avaliar também os resultados das variáveis *escolarização* e *estada fora da comunidade*.

3.3.1 Variável *sexo do falante*

Os resultados quantitativos da variável *sexo* relevam que os homens tendem a empregar mais a regra de concordância verbal que as mulheres na comunidade de Almojarife. Na Tabela 8, verifica-se PR .566 e frequência de 80% de aplicação da regra pelos homens, enquanto as mulheres perfizeram PR .329 e frequência de 56,5%:

Tabela 8. Efeito da variável *sexo* sobre a variável dependente.

Sexo	Nº de oco./Total	Freq.	RODADA 1 P.R.
Masculino	316/395	80%	.576
Feminino	83/147	56%	.306
TOTAL	399/542	73,6%	

Nível de significância: 0.000

Os resultados mostram que os homens favorecem moderadamente a aplicação da regra. Um resultado semelhante é encontrado no estudo de Lucchesi (2009, p. 369) em comunidades afro-brasileiras do interior da Bahia, os homens apresentaram o PR .56 e contra PR .45 das mulheres, numa frequência de 26% e 12% respectivamente.

Porém, na comunidade de São Miguel dos Pretos, no interior do Rio Grande do Sul, Almeida (2006, p. 94)¹² encontrou um resultado oposto. Embora não tenha apresentado os

¹² Em Almeida (2006), o VARBRUL não selecionou a variável *sexo* como relevante, por isso a autora não apresentou os pesos relativos à essa variável.

pesos relativos nem realizado um CROSSTAB para compreender melhor o resultado, Almeida (2006) mostrou a diferença percentual de 75% de aplicação da CV para as mulheres e 72% para os homens. Os exemplos (30) e (31) mostram a aplicação da CV-P4 pelos homens no PA.

(30) *nós prendemos brancos trabalhá* [ABILH2]

(31) *nós conseguimos o campeonato* [CASTEH1]

A maioria dos estudos já realizados sobre as comunidades rurais no Brasil mostra que os homens tendem a liderar as mudanças em favor do padrão normativo porque mantêm mais contatos fora da comunidade do que as mulheres. Geralmente, são os homens que saem para as atividades em busca de trabalho, estão mais envolvidos com a comercialização dos produtos agrícolas nos centros urbanos. Enquanto as mulheres tendem a ficar mais restritas às atividades domésticas e ao trabalho na roça, logo, são os homens que lideram o processo de assimilação dos padrões linguísticos urbanos (Lucchesi, 2009, p.369).

Em Almojarife, muitos homens tiveram contato com o exército colonial e/ou trabalharam nas grandes roças, logo, isso explicaria o fato deles estarem aplicando mais a regra de CV. Almeida (2006, p. 94) explica o resultado oposto à essa proposição devido ao engajamento das mulheres da comunidade de São Miguel dos Pretos com a educação dos filhos (acompanhamento das tarefas escolares) e nas lutas comunitárias. Mas ressalta que somente um aprofundamento nas questões de interação com os falantes e com informações complementares poderia se obter uma justificativa melhor esse resultado.

Para melhor compreensão dos resultados do presente trabalho, foi feito em CROSSTAB um cruzamento da variável *sexo* com a *faixa etária*. Verificou-se que, na faixa 3, as mulheres marcam com mais frequência a CV-P4 do que os homens. E isso acontece apesar da elevada taxa de analfabetismo nessa faixa etária (apenas dois informantes tendo frequentado o ensino primário – um do sexo masculino (3ª. classe) e outro do sexo feminino (1ª. classe), mas sem o concluir). Esse perfil se justificaria porque os homens da faixa etária 3 estão mais envolvidos na pesca e as mulheres com a venda dos produtos. Na faixa 2, a situação se inverte devido à presença da escolaridade e o fato de alguns homens já terem trabalhado no exército colonial e/ou terem trabalhado nas grandes roças de café e cacau, também na época colonial.

3.3.2 Variável *faixa etária do falante*

Foi efetuada uma terceira rodada (R3), incluindo as três faixas etárias e a variável *escolarização* (essa sem a faixa etária 1), mas excluindo a variável *estada fora da comunidade*, que revelou forte interferência nos pesos relativos das demais variáveis ao longo do processo STEPUP e STEPDOWN. Na nova rodada, a *faixa etária* foi selecionada para inclusão no melhor conjunto de variáveis condicionantes. A Tabela 9 apresenta os resultados.

Tabela 9. Efeito da variável *faixa etária* sobre a variável dependente.

Faixa etária do falante	Nº de oco./Total	Freq.	RODADA 3 P.R.
Faixa 1 - 20 a 40 anos	211/247	85,4%	.658
Faixa 2 - 41 a 60 anos	120/180	66,7%	.396
Faixa 3 - Mais de 60 anos	68/115	59,1%	.322
TOTAL	399/542	73,6%	

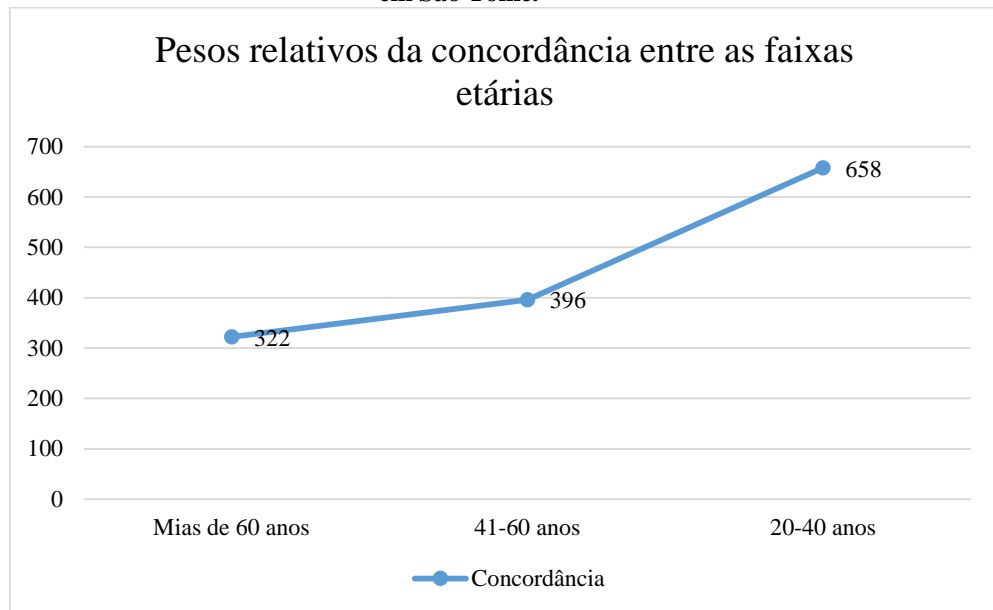
Nível de significância: 0.000

Para diagnosticar qual o processo de variação sincrônica presente numa determinada comunidade de fala, nos termos da dicotomia *variação estável versus mudança em curso*, no tipo de abordagem denominada de *estudo em tempo aparente*, é necessário, primeiramente, observar a distribuição geracional das frequências de uso das variantes na comunidade (LUCCHESI, 2001, p.135; LABOV, 2008 [1972], p. 318). Sendo assim, a variável *faixa etária* é fundamental nessa questão.

Além de confirmar a hipótese levantada de que os mais jovens tendem a aplicar a regra de concordância do que os falantes mais velhos. Os resultados dessa variável também indicam um processo de mudança em curso, pois revelam uma grande inclinação no sentido do aumento do uso da regra de CV.

A análise revelou um quadro de mudança em curso, onde se vê que os falantes da faixa 1 – 20 a 40 anos – são os que lideram na aplicação da regra de CV-P4 com PR .658 e uma frequência de 85,4%. De maneira oposta, os falantes das faixas etárias 2 e 3 apresentam PR progressivamente mais desfavorecedores (faixa 2: .396; faixa 3: .322) e frequências de aplicação progressivamente mais baixas (faixa 2: 66,7%; e faixa 3: 59, 1%), e Esse perfil é visualizado no Gráfico 3.

Gráfico 3. Inclinação dos pesos relativos da CV-P4 entre as faixas etárias na comunidade de Almozarife em São Tomé.



No estudo de Lucchesi (2009), os resultados da variável faixa etária nas comunidades analisadas revelam uma ligeira inclinação no sentido do aumento do uso da regra da CV, na medida em que se passa dos falantes mais velhos para os falantes mais jovens. Porém, a diferença percentual entre as faixas etárias foi mínima, e a variável não foi selecionada pelo VARBRUL como estatisticamente significativa: faixa 1 (20-40 anos) 19%, faixa 2 (41-60 anos) 18%, faixa 3 (acima de 60 anos) 16%. (Lucchesi, 2009, p. 368).

Em Almeida (2006, p. 92) o VARBRUL não selecionou a variável faixa etária, portanto, a autora apresentou apenas os percentuais de frequência encontrados: faixa 1 (15-24 anos) 77%, faixa 2 (40-64 anos) 79% e faixa 3 (65-90 anos) 66%. Esse resultado releva que os mais jovens e os adultos empregam mais a CV-P4 do que os mais velhos. Apesar de não apresentarem PR, os resultados desses dois estudos são semelhantes aos achados neste trabalho. Lembra-se o que já foi mencionado no início deste capítulo: ao cruzar as variáveis *faixa etária* e *escolarização*, detectou-se que não há analfabetos na faixa 1. Isso implica diretamente no fato de os falantes de 20 a 40 anos aplicarem de maneira elevada a regra da CV-P4, como mostram os exemplos (32) e (33).

(32) *nó só íamos pa roça* [DULCEM1]

(33) *Nós fomos a pesca* [OSVALH1]

Outro resultado interessante foi verificado com o cruzamento da *faixa etária* com a *estada fora da comunidade*, em que pode-se notar que nenhum dos informantes da faixa 1 (20 a 40 anos) viveu pelo menos seis meses fora da comunidade. Questões socio-históricas podem explicar isso. Até a independência, em 1975, a economia são-tomense tinha como base a cultura do cacau. Durante os quinze anos subsequentes à independência, essa cultura de cacau se findou e hoje em dia, a economia de São Tomé e Príncipe não tem muito a oferecer nos mercados de exportação e possui uma significativa fragilidade socioeconômica (SANGUIN, 2014, p. 18). A frágil economia do país não oferece muitas oportunidades que façam a população da zona rural ter o anseio de sair das comunidades.

Dessa maneira, entende-se que os informantes almojarifanos de 20 a 40 anos não deixam sua comunidade porque não veem muito proveito nisso. Eles são mais envolvidos nas agricultura de subsistência e na pesca. Nos relatos das entrevistas foi possível notar que há um desejo de permanência na comunidade, bem como de que o crescimento econômico pessoal seja na própria comunidade e não fora dela. Outrossim, relatam a vontade de que o governo invista na comunidade, como se pode ver no relato de [ZECAH1], um informante homem de 25 anos, pescador.

DOC: *Mas você acha que o bom aqui, ocê acha que o bom aqui seria tê uma.... ?*
 INF: *Uma uma loja, pelo memo quê pa sinja, hum tê bucado bucado cola nele ININT gen nã pode compra nada.*
 DOC: *Mas nã tem loja aqui?*
 INF: *Hum?*
 DOC: *Nã tem loja aqui?*
 INF: *Tinha uma loja lá mas tudo seco,*
 DOC: *Hum*
 INF: *Dono saice foi mbora pa Santana nã tem nada*
 DOC: *Hum*
 INF: *E eu tentê de falá com um mê cunhado um que tem um carro la branco*
 DOC: *Hum.*
 INF: *Mê cunhado quê pa impresta- me dinheiro quê pa fazê o hum negócio, ele*
 DOC: *Hum.*
 INF: *É porisso se nã eu podia tê um bucadinho coesa aqui pa vendê,*
 DOC: *Hum.*
 INF: *Pelo memo quê pa vê se vida avança um bocadinho porque só polvo nã dá nada.*
 DOC: *Hum.*
 INF: *E polvo é trabalho muito sforçado e pode dá cabo memo pissoa e eu quero deixa memo de fazê pesca enquanto se eu nã arranja barba barbatana uma arma ou, uma coesa eu quê dexâ memo de fazê pesca porque ê é prigoso e gen pode perdê a vida.*
 DOC: *Hum.*
 INF: *Mas já com mê negócio aqui, ININT mas folgado*
 DOC: *Mas você podia fazê as 2 coesa né?*
 INF: *Hum?*
 DOC: *Podia fazê as 2 coisa, pescava e depois tomava conta do negócio é?*
 INF: *També pode ou ê posso dexa mulhé*
 DOC: *É*
 INF: *Com negócio i fazê pesca quê pa roupa peixe quê pa comê.*
 DOC: *Hum.*

INF: *Uma coesa qualquer, Mas aqui São Tomé gen nã vê ninguém que pa ajudá gente. Pa judá ninguém.*

DOC: *Hum.*

INF: *Ninguém só nã ajuda.*

DOC: *Governo nã dá apoio nã?*

INF: *E, eis tão pecurá só pa bolso deles. Nã ta interessa de nós piqueno é por isso por porquê que há muitos ladrões agora? As vez roba roba bastante é devido isso. As vez tira trabalho as vez uma pissoa vai trabalhar chega mês toma 40 50 conto 1 mês, quê conto faz?*

3.3.3 Variável escolarização

Após ser constatado que não há analfabetos entre os falantes mais jovens, a faixa etária 1, foi feita uma rodada em que os dados desses falantes foram excluídos das variáveis independentes *escolarização* e *estadas fora da comunidade*. Os resultados estão apresentados na Tabela 10.

Tabela 10. Efeito da variável *escolarização* sobre a variável dependente.

Grau de escolaridade	Nº oco./Total	Freq.	RODADA 2 P.R.
Semianalfabeto	145/168	86,3%	.746
Analfabeto	43/127	33,9%	.193
TOTAL	188/295	63%	

Nível de significância: 0.000

Os resultados confirmam a hipótese de que quanto maior o grau de escolaridade, maior também é a tendência de aplicação da regra de CV, conforme os exemplos (34) e (35). Os falantes semianalfabetos aplicaram a regra com PR .746 e uma frequência de 86,3%. Já os falantes analfabetos aplicaram a regra com PR .193 e 33,9% de frequência. Nos estudos de Lucchesi (2009) e Almeida (2006) essa variável não foi comentada por não ser relevante para ambos os trabalhos. Porém, é frequente entre os estudos sobre a CV em geral a observação de que os falantes de maior escolarização realizam mais a concordância.

(34) *é que nós stamos aqui a viver* [CLOTIM2]

(35) *nós pusemos ele na canoa* [ZECAH1]

Em um outro estudo que trabalhou com dados do PST, analisando e comparando a CV – P6 (BRANDÃO E VIEIRA, 2012, p. 22), no PST e no PB (em variedades urbanas), ficou

evidente que quanto maior o nível de escolaridade do indivíduo, menor é a probabilidade de cancelamento da marca de número. No PST, a diferença de peso relativo entre os menos escolarizados (PR .90; com uma frequência de 14,8%) e os mais escolarizados (PR .15; com uma frequência de 1,5%) é expressiva: 75. E, no PB, a situação é semelhante, pois há uma diferença notável entre o desempenho de indivíduos com nível superior (.29; 9,9 %) e o dos de nível fundamental (.61; 23,7 %) e médio (.69; 32%).

O resultado para a variável *escolarização* no PA também permite identificar uma tendência para a aplicação da regra, visto que se não há analfabetos na faixa etária jovem, a instrução escolar teria feito com que os falantes realizem mais a CV. Tal hipótese é corroborada quando se inspeciona o CROSSTAB de *faixa etária* com *escolarização*, pelo resultado dos falantes semianalfabetos da faixa etária intermediária (88%) e, também na faixa etária avançada (79%), entre os falantes que frequentaram até certo período a escola, conforme a Tabela 11:

Tabela 11. Cruzamento de faixa etária com escolarização.

ESCOLARIZAÇÃO	SEMIANALFABETO		ANALFABETO		TOTAL		TOTAL GERAL Nº de ocor.
	Pres. Nº %	Aus. Nº %	Pres. Nº %	Aus. Nº %	Pres. Nº %	Aus. Nº %	
FAIXA ETÁRIA							
Faixa 1 - 20 a 40 anos	211 85%	36 15%	0 ---	0 ---	211 85%	36 15%	247
Faixa 2 - 41 a 60 anos	114 88%	15 12%	6 12%	45 88%	120 67%	60 33%	180
Faixa 3 – Mais de 60 anos	31 79%	8 21%	37 49%	39 51%	68 59%	47 41%	115
TOTAL	356 86%	59 14%	43 34%	84 66%	399 74%	143 26%	542

3.3.4 Variável *estada fora da comunidade*

A Tabela 12 apresenta os resultados da rodada 2 para a variável *estada fora da comunidade*.

Tabela 12. Efeito da variável *estada fora da comunidade* sobre a variável dependente.

Tempo da estada fora	Nº oco./Total	Freq.	RODADA 2 P.R.
Sempre viveu na comunidade	53/122	43,4%	.290
Viveu pelo menos 6 meses fora	135/173	78%	.653
TOTAL	188/295	63,7%	

Nível de significância: 0.000

Os informantes que viveram pelo menos seis meses fora da comunidade registraram PR .653 e percentual de 78% de aplicação da CV-P4. Em contrapartida, os informantes que sempre viveram na comunidade registraram PR .290 e frequência de 43,4% de concordância. Dessa maneira, confirmou-se a hipótese de que os falantes que viveram determinado tempo fora da comunidade marcariam mais a CV. Esse fato evidencia uma tendência para mudança aquisicional de cima para baixo, já que mostra a influência do contato no exército colonial e do contato no âmbito do trabalho nas grandes roças.

(36) *nós aqui criamos é deus só sabe* [MMDEUSM3]

(37) *nós jogamos bola* [MAURIH2]

Lucchesi (2009, p. 369-370) obteve resultados semelhantes a estes para as comunidades afro-brasileiras do interior da Bahia. Ele constatou que aqueles indivíduos que viveram fora das comunidades por um período superior a seis meses (geralmente trabalhando no Sul do Brasil) são os que mais empregam a regra de CV-P4, com PR .59 e percentual de 23% de frequência, contra PR .40 e 12% de frequência daqueles que sempre viveram nas comunidades. Nessa base, afirmou-se que “Essa seria então a evidência mais forte em favor de uma mudança de fora para dentro da comunidade em função de um processo de nivelamento linguístico que se expande a partir dos grandes centros urbanos.” (LUCCHESI, 2009, p. 370).

3.4 SÍNTESE

Neste capítulo foram apresentados os resultados da análise dos dados da CV-P4. Na seção foram mostrados os resultados gerais: 542 ocorrências da variável dependente, sendo 399 ocorrências com concordância, correspondendo a 73,6% dos dados, e 143 sem concordância, totalizando 26,4%. O percentual dos dados com concordância na fala da comunidade de Almojarife foi inverso ao que se esperava, pois a suposição inicial era que se encontraria mais ausência. No entanto, os dados mostram que a concordância faz parte da gramática dessa comunidade.

A quantificação preliminar dos dados revelou a presença de KNOCKOUTS, os quais foram eliminados. Para verificar melhor os resultados, foi feito o cruzamento de cada par de variáveis independentes por meio de CROSSTAB, o qual revelou lacunas na distribuição dos dados. Com isso, foram eliminados algumas variáveis e logo após, foram realizadas três rodadas de VARBRUL.

Os resultados das três rodadas foram apresentados nas seções 3.2 e 3.3. Na seção 3.2 foram expostas as variáveis referentes ao encaixamento estrutural da CV-P4: **realização e posição do sujeito e saliência fônica**, selecionadas nessa mesma ordem pelo VARBRUL. Apresentou-se as frequências da variável **paralelismo formal**, mesmo sem esta ter sido avaliada pelo programa, pois o percentual dessa variável foi interessante.

O encaixamento social foi mostrado na seção 3.3, na qual foram expostos os resultados das variáveis: **sexo, faixa etária, escolarização e estada fora da comunidade**. A comparação dos resultados do PA com outras sobre o PB relevou uma similaridade entre essas variedades do português, o que leva considerar que o português reestruturado, adquirido em situação de contato por falantes de línguas crioulas tem os mesmos padrões de variação do português adquirido numa situação de transmissão linguística irregular na qual a L1 ancestral não é um crioulo, como o PB.

CONSIDERAÇÕES FINAIS

Neste trabalho investigou-se o quadro de variação da CV-P4 na comunidade de Almojarife, em São Tomé, a qual é formada por falantes bilíngues em crioulo santome e em português. Com o intuito de contribuir para a discussão sobre o português africano, uma das questões levantadas foi verificar se a realidade linguística do PA apresenta características semelhantes ou distintas da variedade brasileira do português. Para tanto, recorreu-se à comparação de estudos já feitos sobre o mesmo tema para o PB.

Para averiguar essas propostas, foram analisados dados da fala de 18 informantes de ambos os sexos, divididos em três faixas etárias. A análise dos dados partiu da hipótese geral de que a variação no PA seria motivada por fatores linguísticos e extralinguísticos que refletem características tipológicas, atuais e ancestrais, da L1 dos aprendentes. Bem como, apresenta aspectos universais, isto, é independentes da L1 e da L2, que atuaram na estrutura do português quando este era L2 e, conseqüentemente, passou a atuar no português quando se tornou L1 nativizada.

A pesquisa apoiou-se no modelo da Sociolinguística Laboviana (Labov, 1972) e recorreu ao Goldvarb (2001), programa computacional que abrange o pacote de dados do VARBRUL, para a quantificação e análise probabilística dos dados referente à CV-P4. Das 542 ocorrências do fenômeno encontradas no *corpus* estudado, 399 foram com concordância, equivalendo a 73,6% dos dados, e 143 sem concordância, equivalendo a 26,4%.

As variáveis independentes inicialmente propostas para esta pesquisa foram: **realização e posição do sujeito, nível de referencialidade, paralelismo formal, saliência fônica, faixa etária, sexo, escolarização e estada fora da comunidade**. Porém, no processamento estatístico, foram selecionadas como significativas na primeira rodada geral: **realização e posição do sujeito, saliência fônica, faixa etária e sexo**. Nas rodadas adicionais, que excluíram das variáveis sociais os dados da faixa etária 1 (20-40 anos), foram selecionadas: **escolarização e estada fora da comunidade**.

Os resultados obtidos permitiram identificar que:

(i) quando o sujeito não foi realizado a aplicação da CV foi favorecida com um PR .657. Sendo confirmada a hipótese de que a CV seria mais realizada quando não houvesse a realização fonética do sujeito. Prevalece uma lógica funcional, pois o morfema de número e pessoa supriria a ausência de informação com o sujeito nulo;

(ii) na saliência fônica, foi possível verificar que os níveis 2 (*fala/falamos*), 4 (*vai/vamos*) e 5 (*é/somos*) são os que favorecem a concordância verbal, apresentando,

respectivamente, PR de .597, .531, e .567. Os favorecimentos relativos dos níveis 4 e 5 já eram esperados, e, portanto, corroboram a hipótese levantada – de que os verbos mais salientes favorecem a CV. Porém o PR do nível 2, com um favorecimento moderado de .597, não era esperado, sendo que o nível 2 em estudos prévios sobre a CV-P4 corresponde a um grau de saliência não proeminente. Foi proposto que esse resultado se deve ao fato de que no nível 2 da escala a sílaba tônica muda de posição quando o verbo passa do singular para o plural, como em *fala/falamos*. Sendo os verbos no crioulo uniformes, o falante tende a marcar a diferenciação que é mais perceptível.

(iii) A variável paralelismo discursivo não foi processada no VARBRUL por apresentar, por um lado, alguns fatores categóricos e com poucos dados, e por outro lado, fatores que não apresentavam uma distribuição ortogonal com as outras variáveis independentes. Contudo, a frequência geral dos fatores pode ser sugestiva. Quando o verbo é precedido por *nós + verbo com desinência – mos/-mo, na oração anterior* o percentual de concordância foi de 99,6%. A frequência da forma verbal precedida por *verbo com desinência – mos/-mo, sem sujeito realizado, na oração anterior* também foi alta, 88,2%. E quando forma verbal era *isolada ou primeira referência* apresentou a frequência de 86% de favorecimento à CV;

(iv) Em contraste com o desempenho das mulheres, os homens tendem a empregar mais a regra de CV-P4, com um peso relativo de .566. Esse fato é explicado porque os homens mantêm mais contatos fora da comunidade do que as mulheres. No caso de Almojarife, o contato com o exército colonial e o trabalho nas grandes roças pode ter influenciado a aquisição de marcas do padrão da LA;

(v) Os falantes mais jovens – 20 a 40 anos – são os que lideram na aplicação da regra de CV-P4 com um PR .685 e. Confirmou-se a hipótese de que, contrariamente ao desempenho dos falantes mais velhos, os mais jovens tenderiam a aplicar a regra de concordância;

As variáveis que não foram selecionadas pelo VARBRUL apresentaram resultados interessantes quando tabuladas em cruzamento com outras variáveis. Ao efetuar uma tabulação cruzada da variável *faixa etária* com a *escolarização*, por meio do programa CROSSTAB, detectou-se que não há analfabetos na faixa 1. Isso implica diretamente no fato de os falantes de 20 a 40 anos aplicarem de maneira elevada a regra da CV-P4.

Outra tabulação cruzada efetuada por meio de CROSSTAB avaliou a variável *sexo* e a *faixa etária*. Notou-se que, na faixa 3, as mulheres marcam mais a CV-P4 do que os homens,

mesmo sendo elevada a taxa de analfabetismo dessa faixa etária. Isso se justificaria porque os homens dessa faixa etária estão mais ligados à pesca e as mulheres com a venda dos produtos.

Os resultados ainda mostraram que:

(vi) Os falantes que possuíam algum grau de escolarização foram os que mais aplicaram a CV-P4, permitindo identificar uma tendência para a aplicação da regra;

(vii) Os informantes que viveram por, pelo menos, seis meses fora da comunidade empregaram mais a CV do que os que sempre viveram na comunidade.

Constatou-se que a variação da CV-P4 é orientada por variáveis linguísticas que se revelaram ser importantes em estudos prévios sobre o PB, com exceção do comportamento da saliência fônica. Estes aspectos permitem detectar os tipos de estruturas em que os aprendentes do português L2 que têm a língua crioula santome como L1 revelam propensão para realizar, ou não, as marcas formais de plural.

No que tange ao encaixamento social da CV-P4 na comunidade de Almojarife, os resultados indicam um processo de mudança em curso, pois revelam uma grande inclinação no sentido do aumento do uso da regra de CV. Verifica-se esse fato através dos PR da variável faixa etária: faixa 1 .658, faixa 2 .396 e faixa 3 .322. A comparação dos resultados de cada variável e do resultado geral com os estudos (ALMEIDA, 2006; LUCCHESI, 2009) acerca do PB, permitiu averiguar que o perfil da variação na CV-P4 no PA é semelhante ao do PB, visto que o comportamento das variáveis estruturais e sociais desses estudos são semelhantes.

Verificou-se também que os falantes adultos aprendentes de L2 acessam parcialmente à GU, isto é, os parâmetros desativados após o período crítico da aquisição não são totalmente disponibilizados. Isso fica evidente no resultado da variável faixa etária, onde se vê que os mais velhos apresentam um índice menor de aplicação da CV, a gramática dos falantes mais idosos evidenciando proximidade à gramática do substrato e a das gerações mais novas caminhando no sentido de uma ligeira aquisição das regras da gramática da LA, ou seja, há gramáticas em competição.

A partir disso, conclui-se que o português L2 e posteriormente L1, adquirido em situação de contato por falantes de línguas crioulas apresenta os mesmos padrões de variação do português adquirido numa situação de transmissão linguística irregular na qual a L1 ancestral não é um crioulo, situação semelhante ao PB, conforme já observara Figueiredo (2010).

Nesta conformidade, a explicação para a formação das línguas reestruturadas, para além de levar em conta as questões relacionadas com o tipo de aquisição e o desenvolvimento da L2 e L1 até ao estágio final de ILn ou língua-I, não pode também descartar a matriz sociolinguístico-histórica em que a aquisição se dá (HOLM, 1998; 2004), assimilando as perspectivas da transmissão linguística irregular, com influências das L1's ancestrais (WINFORD, 2003a) e das línguas pidgin ou línguas crioulas. (FIGUEIREDO, 2010, p.704).

Portanto, na estrutura do PA, que é uma variedade reestruturada do português, agem paralelamente fatores relacionados à aquisição da linguagem e aspectos sociolinguísticos e históricos em que a aquisição ocorreu. Respondidas às questões que nortearam e motivaram o presente trabalho, destacam-se algumas limitações deste trabalho.

Acredita-se que com um *corpus* com maior número de informantes será possível verificar a funcionalidade de variáveis que não foram selecionadas neste estudo devido ao pouco número de dados em relação à elas, apesar do fato das limitações demográficas da comunidade dificultarem a configuração de um *corpus* maior.

Outra questão relevante é verificar o papel da escolaridade na aquisição da variedade padrão, já que essa variável apresentou um resultado interessante entre os falantes mais jovens. A inserção do pronome *a gente* também poderia acrescentar informações sobre a variação da CV-P4 que não foram elencadas nesta dissertação.

Por fim, afirma-se que este estudo trouxe uma contribuição ao desvendamento da realidade sociolinguística do PA, e em certa forma do PB ao confirmar que ambos são similares.

REFERÊNCIAS

ALMEIDA, A. P. 2006. **A concordância verbal na comunidade de São Miguel dos Pretos, Restinga Seca, RS**. Dissertação (Mestrado em Estudos da Linguagem) - Universidade Federal do Rio Grande do Sul, Porto Alegre.

AO, S. S. H. 2013. **Aquisição da morfologia verbal do português europeu: um caso de aprendentes de chinês-L1**. Faculdade de Ciências Sociais e Humanas. Tese de doutorado. Universidade de Macau.

BACCEGA, M. A. 1989. **Concordância Verbal**. 2ª edição. São Paulo: Ática.

BAXTER, A. N. 1992. A contribuição das comunidades afro-brasileiras isoladas para o debate sobre a criouliização prévia: um exemplo do estado da Bahia. In: E. ANDRADE, (ed), **Actas do Colóquio Internacional sobre Línguas Crioulas de Base Portuguesa**, Universidade de Lisboa. Lisboa: Colibri, p.7-35.

BAXTER, A. N. 1995. Transmissão Geracional Irregular na História do Português Brasileiro - divergências nas vertentes afro-brasileiras. **Revista Internacional de Língua Portuguesa** 14:72-90.

BAXTER, Alan Norman. 2004. The development of variable NP plural agreement in a restructured african variety of portuguese. In: ESCURE, G.; SCHWEGLER, A. (eds.), **Creoles, contact, and language change: linguistic and social implications**, vol. 27, 97-126. Amsterdam: John Benjamins.

BAXTER, Alan Norman. 2009. A concordância de número. In: LUCCHESI, D.; BAXTER, A.; RIBEIRO, I.(orgs.). **O português afrobrasileiro**, 269-293. Salvador: EDUFBA.

BAXTER, A. N.; LUCCHESI, D. 1997. A Relevância dos Processos de Pidginização e Crioulização na Formação da Língua Portuguesa. In: **Estudos Linguísticos e Literários**. Salvador. Universidade Federal da Bahia, Programa de Pós-Graduação em Letras e Linguística, n.19, p. 65-84.

BICKERTON, D. How to acquire language without positive evidence: what acquisitionists can learn from Creoles? In: DEGRAFF, M. (ed.). **Language creation and language change: creolization, diachrony, and development**. Cambridge: The MIT Press, 1999. p.49-74.

BRANDÃO, S. F.; VIEIRA, S. R. 2012. A concordância nominal e verbal no Português do Brasil e no Português de São Tomé: uma abordagem sociolinguística. **PAPIA** 22(1), p. 7-39.

CHAMBERS, J. K.; TRUDGILL, P. 1980. **Dialectology**. Cambridge: Cambridge University Press.

CHOMSKY, N. 1981. **Lectures on government and binding**. Dordrecht: Foris.

ELLIS, R. 1997. **Second language acquisition**. Hong Kong: Oxford University Press.

FIGUEIREDO, C. F. G.. 2008. A concordância variável no sintagma nominal plural do português reestruturado de Almojarife (São Tomé). *PAPIA*, 18. 23-43.

FIGUEIREDO, C. F. G. 2009. A configuração do SN do português reestruturado da comunidade de Almojarife – São Tomé [em linha]. *RCBLPE – Revista de Crioulos de Base Lexical Portuguesa e Espanhola*, 1(1). 28-55.. Disponível em: <http://rcblpe.sftw.umac.mo/doc/A20ConfiguraC3A7ao20do20SN20Plural20do20PortugC3Restruturado20da20Comunidade20de20Almojarife20-20Sao20TomC3A9.pdf>. Acesso em 17. Jan. 2015.

FIGUEIREDO, C. F. G. 2010. **A concordância plural variável no sintagma nominal do português reestruturado da comunidade de Almojarife, São Tomé** (Desenvolvimento das regras de concordância variáveis no processo de transmissão-aquisição geracional). Tese de Doutorado, Universidade de Macau.

GONÇALVES, P. 2000. A gênese de línguas formadas em contextos multilingues: uma abordagem paramétrica. In: ANDRADE, E.; MOTA, M. A. C. da.; PEREIRA, D. (orgs.), **Crioulos de base portuguesa (Actas do workshop sobre crioulos de base lexical portuguesa. FLUL, 29 e 30 de Abril de 1999)**, 247-257. Braga: Associação Portuguesa de Linguística.

GONÇALVES, R. M. G. 2010. **Propriedades de subcategorização verbal no português de S. Tomé**. Dissertação de Mestrado em Linguística, Universidade de Lisboa.

GUY, G. R. 1981. **Linguistic Variation in Brazilian portuguese**: aspects of the Phonology, Syntax, and Language History. Ph.D. Dissertation, University of Pennsylvania.

GUY, G. R.; ZILLES, A. 2007. **Sociolingüística quantitativa** – Instrumental de análise. São Paulo: Parábola Editorial.

HAGEMEIJER, T. 1999. As ilhas de babel: a criouliização no golfo da guiné. *Revista Camões*. VI, 74-88.

HAGEMEIJER, T. 2009. As línguas de São Tomé e Príncipe. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* 1, 1: 1-27.

HOLM, J. 1998. **Pidgins and Creoles**. Cambridge: Cambridge University Press.

KENEDY, E. 2013. **Curso básico de linguística gerativa**. São Paulo: Contexto.

KLEIN, W. 1986. **Second language acquisition**. Cambridge: Cambridge University Press, 1986.

LABOV, W. 1972. **Sociolinguistic Patterns**. Philadelphia: University of Pennsylvania Press. [Padrões Sociolingüísticos. Trad.: Marcos Bagno; Marta Scherre e Caroline Cardoso. São Paulo: Parábola, 2008.]

LABOV, W. 1982. Building on Empirical Foundations. In: LEHMANN, W.; MALKIEL, Y. (orgs.) **Perspectives on Historical Linguistics**. Amsterdam: John Benjamins: 17-92.

- LABOV, W. 1994. **Principles of Linguistic Change**. Oxford/Cambridge: Blackwell.
- LABOV, W. 2001. **Principles of Linguistic change**. Volume II: Social Factors. Oxford: Blackwell.
- LEFEBVRE, C. 1998. **Creole genesis and the acquisition of grammar**: the case of Haitian Creole. Cambridge: Cambridge University Press.
- LEFEBVRE, C. 1997. Relexification in creole genesis: the case of demonstrative terms. In: Haitian creole. **Journal of pidgin and creole languages** 12:2.181-201.
- LIGHTFOOT, D. 1989. The child's trigger experience: degree-0 learnability. In: **Behaviorial and Brain Sciences**, 12(2). Cambridge, MA: Cambridge University
- LOPES, N.; BAXTER, A. N. 2009. **O desenvolvimento de regras de concordância variável em variedades de Português a partir de modelos-estímulos (inputs) diferentes**. ROSAE - I Congresso Internacional de Linguística Histórica, Salvador, Bahia, Brasil, 26-29 Julho.
- LOPES, N.; BAXTER, A. N. 2011. A concordância verbal variável no português dos Tongas. **PAPIA** 21 (1), 39-50.
- LUCCHESI, D. 1998. **Sistema, mudança e linguagem**. Lisboa: Colibri.
- LUCCHESI, D. 2000. **Reanálise da variação na concordância de gênero em um dialeto afrobrasileiro**. XVIII Jornada de Estudos Linguísticos do GELNE. Salvador: Universidade Federal da Bahia.
- LUCCHESI, D. 2001. O tempo aparente e as variáveis sociais. **Boletim da ABRALIN**, v.26, p.135-137, Número especial.
- LUCCHESI, D. 2004. **Sistema, Mudança e Linguagem**. São Paulo: Parábola.
- LUCCHESI, D.; ARAUJO, S. S. de F. 2004. **A teoria da variação linguística**. Disponível em: <<http://www.vertentes.ufba.br/a-teoria-da-variacao-linguistica>>. Acesso em 23. março. 2015.
- LUCCHESI, D. 2008. A concordância nominal em estruturas passivas e de predicativo do sujeito em comunidades rurais afro-brasileiras isoladas no contexto da história sociolingüística do Brasil. In: VOTRE, S.; RONCARATI, C. (orgs.). **Anthony Julius Naro e a lingüística no Brasil: uma homenagem acadêmica**, 1ª. ed., vol. 1, 148-168. Rio de Janeiro: 7 Letras.
- LUCCHESI, D. 2009. A concordância verbal com a primeira pessoa do plural. In: LUCCHESI, D.; BAXTER, A. N.; RIBEIRO, I. (orgs). **O português afro-brasileiro**. 360-371. Salvador: EDUFBA.
- LUCCHESI, D.; BAXTER, A. N. 2009. A transmissão linguística irregular. In: LUCCHESI, D.; BAXTER, A. N.; RIBEIRO, I. (orgs). **O português afro-brasileiro**. 101-124. Salvador: EDUFBA.

- MOLLICA, C. 2003. Fundamentação teórica: conceituação e delimitação. In: MOLLICA, C.; BRAGA, M. L. (orgs.). **Introdução à Sociolinguística: o tratamento da variação**. São Paulo: Contexto, p. 9-14.
- MORENO, A.; TUZINE, A. 1998. Distribuição social de variáveis linguísticas no português oral de Maputo. In: STROUD, C.; GONÇALVES, P. (orgs.), **Panorama do português oral de Maputo. Volume II: a construção de um banco de “erros”**, 68-89. Maputo: Instituto Nacional do Desenvolvimento da Educação.
- NARO, A. 1981. The social and structural dimensions of syntactic change. **Language**, v. 57, n.1, p. 63-98.
- NARO, A. 2003. O dinamismo das línguas. In.: MOLLICA, C.; BRAGA, M. L. (orgs.). **Introdução à Sociolinguística: o tratamento da variação**. São Paulo: Contexto, p. 43-51.
- NARO, A.; LEMLE, M. 1976. Syntactic diffusion. **Ciência e Cultura**, v. 29, n. 3, p.259-268.
- NARO, A.; SCHERRE, M. M. P. 1993. Sobre as origens do português popular do Brasil. **DELTA**, n. 9, p. 437-454.
- NARO, A.; SCHERRE, M. M. P. 2000. Variable concord in portuguese: the situation in Brasil and Portugal. In: Mc WORTHER, J. (Ed.). **Language change and language contact in pidgins and creoles**. Amsterdam/Philadelphia: John Benjamins Publishins, p. 235-255.
- NARO, A.; SCHERRE, M. M. P. 2007. **Origens do português brasileiro**. São Paulo: Parábola.
- RGPH – 1981. (1987). **Recenseamento geral da população e da habitação**. Lisboa: Instituto Nacional de Estatística de São Tomé e Príncipe.
- RGPH – 1991. **II Recenseamento geral da população e da habitação**. República Democrática de S. Tomé e Príncipe.
- RGPH – 2001. (2003). **Características educacional da população**. Instituto Nacional de Estatística de São Tomé e Príncipe.
- SANGUIN, A. L. 2014. São Tomé e Príncipe, as ilhas do meio do mundo: avaliação crítica sobre sua geografia política. **Confins** [Online]. Disponível em: <http://confins.revues.org/9018>; DOI: 10.4000/confins.9018 Acesso em 23 Abril 2015.
- SCHERRE, M. M. P. 1988. **Reanálise da concordância nominal em português**. Tese de Doutorado. Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- SILVA, J. A. da. A concordância verbal com a terceira pessoa do plural. In: LUCCHESI, D.; BAXTER, A. N.; RIBEIRO, I. (orgs.). **O português afrobrasileiro**, 348-360. Salvador: EDUFBA.

TAGLIAMONTE, S. A. 2012. **Variationist Sociolinguistics: change, observation, interpretation.** Oxford: Blackwell.

TARALLO, F. 1986. **A Pesquisa Sociolinguística.** São Paulo: Ática.

WEINREICH, W.; LABOV, W.; HERZOG, M. 1968. Empirical Foundations for Theory of Language Change. In: LEHMANN, P.; MALKIEL, Y. (orgs.) **Directions for Historical Linguistics.** Austin: University of Texas Press: 95-188. [**Fundamentos empíricos para uma teoria da mudança linguística.** Trad.: Marcos Bagno; revisão técnica: Carlos Alberto Faraco. São Paulo: Parábola, 2006.]

WHITE, L. 1989. **Universal Grammar and second language acquisition.** Amsterdam/Philadelphia: John Benjamins.

WINFORD, D. 2003. **An Introduction to contact linguistics.** Oxford: Blackwell.