



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

**A CONCORDÂNCIA NOMINAL DE NÚMERO NO PORTUGUÊS DO  
INTERIOR DA BAHIA: INTERFERÊNCIA DO CONTATO ENTRE  
LÍNGUAS?**

**THAMIRIS SANTANA COELHO ASSIS**

Salvador

2022

**THAMIRIS SANTANA COELHO ASSIS**

**A CONCORDÂNCIA NOMINAL DE NÚMERO NO PORTUGUÊS DO  
INTERIOR DA BAHIA: INTERFERÊNCIA DO CONTATO ENTRE  
LÍNGUAS?**

Tese apresentada ao Programa de Pós-Graduação em Língua e Cultura do Instituto de Letras da Universidade Federal da Bahia – UFBA, como parte dos requisitos para obtenção do grau de Doutora em Língua e Cultura.

Orientadora: Profa. Dra. Maria Cristina Vieira de Figueiredo Silva

Coorientadora: Profa. Dra. Lanuza Lima Santos

Salvador

2022

Dados internacionais de catalogação-na-publicação  
(SIBI/UFBA/Biblioteca Universitária Reitor Macedo Costa)

Assis, Thamiris Santana Coelho.

A concordância nominal de número no português do interior da Bahia: interferência do contato entre línguas? / Thamiris Santana Coelho Assis. - 2022.

209 f.: il.

Orientadora: Profa. Dra. Maria Cristina Vieira de Figueiredo Silva.

Coorientadora: Profa. Dra. Lanuza Lima Santos.

Tese (doutorado) - Universidade Federal da Bahia, Instituto de Letras, Salvador, 2022.

1. Sociolinguística. 2. Contato linguístico. 3. Língua portuguesa - Concordâncias. 4. Língua portuguesa - Português falado - Poções (BA). 5. Língua portuguesa - Português falado - Santo Antônio de Jesus (BA). I. Silva, Maria Cristina Vieira de Figueiredo. II. Universidade Federal da Bahia. Instituto de Letras. III. Título.

*A minha família, meu bem mais  
precioso.*

## AGRADECIMENTOS

“Gratidão é a expressão da minha alma” àqueles que estiveram comigo durante as etapas – que não foram poucas – deste trabalho.

Agradeço imensamente à minha orientadora, Maria Cristina Figueiredo, pelo empenho e dedicação em orientações excelentes durante todo o percurso do doutorado. A ela, minha gratidão pela confiança, pela compreensão, pelo afeto, pelo apoio e por ser uma referência para mim.

A Lanuza Santos, pelas sugestões preciosas antes mesmo de ser tornar minha co-orientadora, pelo encorajamento diante daquilo que se mostrava novo e temido, pelo olhar atento e pelo cuidado com a pesquisa.

Às professoras Lílian Teixeira e Ivana Ivo pelo conhecimento partilhado acerca dos povos africanos e dos povos indígenas, bem como suas línguas. Agradeço a essas professoras pela empatia, humanidade e competência durante o Semestre Suplementar da UFBA, que ocorreu em um período bastante crítico da pandemia.

Aos professores Emerson Souza, João Paulo Cyrino e à professora Lívia Oushiro pelas orientações valiosas acerca da estatística para linguistas, pelos ensinamentos sobre o R e pela disponibilidade em sanar as dúvidas sempre que necessário.

A professora Vívian Antonino pelas valiosas contribuições durante a qualificação deste trabalho.

A professora Simone Bueno pela supervisão de excelência durante o estágio docente na UFBA, no qual pude aprimorar minhas habilidades e competências quanto ao ensino.

Aos meus alunos do estágio docente por terem me proporcionado momentos de intenso aprendizado e de muitas alegrias.

Aos meus alunos das redes municipal e estadual de ensino das cidades de Ipecaetá, Rafael Jambreiro e Santo Estevão, dos quais precisei ausentar-me para dedicação exclusiva ao doutorado. Muito obrigada por todas as demonstrações de carinho e de saudade.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) pelo financiamento do projeto de pesquisa que permitiu a realização deste trabalho.

Ao Projeto Vertentes pela disponibilização do *corpus* e por todo o conhecimento científico compartilhado.

Aos funcionários do PPGLinc pela competência e solicitude.

Ao meu grupo de “estudo e resenha preta” por ser minha rede apoio mais forte nessa jornada, principalmente, nos dois últimos anos, nos quais enfrentamos a pandemia da covid-19 e vivemos momentos muito difíceis em todas as áreas da nossa vida. Eu não teria chegado até aqui sem vocês! Definitivamente, cada momento de discussão linguística, cada encontro virtual para estudarmos, cada dúvida compartilhada, cada dor e cada alegria, construíram uma amizade que vai além da jornada acadêmica.

Obrigada, Amanda Quiroga, pela escuta atenta, pelo carinho em cada conselho, pela ajuda com a língua inglesa, *you are the best!*

Obrigada, Larissa Santana, por cada palavra de incentivo, pelas ricas discussões sobre a sociolinguística e pelo compartilhamento de textos relevantes da nossa área.

Obrigada, Ludquellen Dias, pela parceria em enfrentar o caminho do R, por todo o cuidado e por todo o auxílio durante o trabalho com os dados, pela leitura atenta dos meus textos sempre que solicitei.

Agradeço a Leandro Santos, pelo encorajamento nos momentos difíceis, pela amizade tão bem consolidada, pelos risos contagiantes que tornaram o percurso, sem dúvidas, mais leve.

A Ramon Paranhos, pelas contribuições inteligentes acerca do kimbundo, pela parceria nos trabalhos acadêmicos e pela leveza de sua amizade.

A Adilson Silva, a Isabel Silveira, e a Rose Cunha, pelo companheirismo desde a época do mestrado e por continuarmos juntos para além da UFBA.

Externo a minha gratidão a: Amanda Galiza, Daniela Alves, Eleneide Oliveira, Gracielle Barros, Ísis Barros, Jéssica Carneiro e Raísa Reis pelo compartilhamento angústias e das alegrias da pós-graduação.

A Deyse Edberg, companheira desde a graduação na UEFS, pela trajetória compartilhada até o doutorado, pela amizade e pelo carinho.

Aos meus amigos da vida, que entenderam minha ausência em muitos momentos ao longo desses anos e continuaram sendo porto seguro. A Joelaine Santana, pela amizade de uma vida inteira, por sempre me acolher em tudo que está ao seu alcance.

A Heilaine Suzart pela generosidade, por me ensinar tanto sobre a vida e por não soltar a minha mão nos momentos mais delicados desse percurso e para além dele.

À minha família pelas orações e por todo incentivo.

À minha mãe Adailza Santana, por ser meu maior exemplo de ser humano, pelo amor, pela motivação, por não medir esforços em dar o melhor para seus filhos: educação.

Ao meu pai Renato Coelho pelo incentivo, por me ensinar que não há limites para os sonhos.

Aos meus irmãos Sillas Coelho, Paulo Natã Coelho e Naara Coelho pela torcida constante, pelo orgulho sempre demonstrado e pela nossa união. À Carol Carvalho pelos momentos de diversão, tão necessários para a vida.

Ao meu sobrinho amado Guilherme Coelho, ainda pequenino, cujo amor não consigo descrever em palavras.

A Daniel Assis, pelo amor e carinho dedicados a mim constantemente, e por entender meus momentos de quietude ao estudar.

E ao meu Deus, o meu Senhor Jesus Cristo, pela misericórdia e graça dispensadas sobre mim, pelo seu doce Espírito Santo, o amado de minh' alma. "Porque Dele e por Ele, e para Ele, são todas as coisas; glória, pois, a ele eternamente" (Rm. 11:36).

## RESUMO

Neste estudo, analiso a variação no uso da concordância de número no sintagma nominal (CN) no português popular do interior da Bahia (sede e zona rural) dos Municípios de Poções e Santo Antônio de Jesus, com o objetivo de discutir o *status* sociolinguístico desse fenômeno num *continuum* linguístico (BORTONI-RICARDO, 1998, 2005; LUCCHESI, 2015), em que, de um lado, está o português rural afro-brasileiro baiano (ANDRADE, 2003, 2006; BAXTER, 2009) e, do outro, o português popular urbano de Salvador (LOPES, 2001). Conforme Lucchesi (2000), o fenômeno da CN afeta os dois extremos das normas linguísticas culta e popular, polarizando-as: a aplicação da regra de CN é muito mais favorecida nas comunidades urbanas e desfavorecida na zona rural. A hipótese principal deste trabalho é de que o português do interior da Bahia seja uma variedade intermediária tendo em vista que as comunidades em estudo mantêm contato com essas duas normas linguísticas. Tendo como referencial teórico-metodológico a Sociolinguística Variacionista (LABOV, 2008 [1972]), verifico o comportamento das variáveis linguísticas e extralinguísticas no uso da CN nos *corpus* do português do interior da Bahia, acervo do Projeto Vertentes (Universidade Federal da Bahia). Analisei os dados da fala de 48 informantes dos sexos masculino e feminino, com pouca ou nenhuma escolaridade, distribuídos em três faixas etárias (25-35; 45-55; mais de 65 anos), recorrendo ao tratamento estatístico R (R Core Team, 2021). Os resultados gerais confirmam a hipótese levantada, pois o percentual de aplicação da regra de CN perfez 17,3% dos dados, sendo intermediário aos resultados do português afro-brasileiro e do português de Salvador. Além disso, na zona rural das comunidades em estudo, a aplicação da regra é menos favorecida, comportamento semelhante ao que se verifica no PB de modo geral. Dentre as variáveis linguísticas, destacam-se a configuração sintagmática do SN e o número de constituintes flexionáveis do SN, revelando que a presença de um item com valor semântico de plural favorece a concordância. O resultado da saliência fônica destoa da maioria dos estudos sobre a CN. Quanto à variável saliência fônica, o resultado obtido não favorece a aplicação da regra, contrariando alguns trabalhos sobre o tema (LOPES, 2001; ANDRADE, 2006). A análise estatística das variáveis sociais indica que os homens realizam mais a concordância que as mulheres, e que a escola não possui um papel tão forte como agente de difusão da norma de prestígio. Os resultados da faixa etária indicam uma tendência aquisicional com os falantes mais jovens aplicando mais a regra de concordância do que os falantes mais velhos. Para explicar a correlação entre as características tipológicas do português e das línguas bantu, adoto a perspectiva da evolução das línguas a partir da competição e seleção de traços (MUFWENE, 2002, 2008) e da teoria nula (ABOH, 2015; ABOH e DEGRAFF, 2017). Defendo que, durante o período de colonização do Brasil, teria ocorrido um processo de reanálise gramatical na aquisição do português como segunda língua em que os prefixos das línguas bantu foram reinterpretados seguindo a manutenção inicial da gramática de aglutinação dos elementos funcionais observados nas línguas bantu, o que justificaria a marcação de plural apenas no primeiro elemento do SN.

**Palavras-chave:** concordância nominal; português do interior da Bahia; sociolinguística; contato linguístico.



## ABSTRACT

In this study, I analyze the variation in the use of number agreement in the noun phrase (NA) in popular Portuguese in the inland cities of Bahia (headquarters and rural areas) in the cities of Poções and Santo Antônio de Jesus, aiming the discussion of the sociolinguistic status of this phenomenon in a linguistic continuum (BORTONI-RICARDO, 1998, 2005; LUCCHESI, 2015), which on one side is Bahian Afro-Brazilian rural Portuguese (ANDRADE, 2003, 2006; BAXTER, 2009) and on the other, the urban popular Portuguese of Salvador (LOPES, 2001). According to Lucchesi (2000), the phenomenon of NA affects both extremes of cultured and popular linguistic norms, polarizing them: the application of the NA rule is much more favored in urban communities and disadvantaged in rural areas. The main hypothesis of this work is that Portuguese from the inland cities of Bahia is an intermediate variety, considering that the communities under study maintain contact with these two linguistic norms. Using Variationist Sociolinguistics (LABOV, 2008 [1972]) as a theoretical-methodological referential, I verify the behavior of linguistic and extralinguistic variables in the use of NA in the corpus of Portuguese in the inland cities of Bahia, a collection of the Projeto Vertentes (Federal University of Bahia). I analyzed the speech data of 48 male and female informants, with little or no schooling, distributed in three age groups (25-35; 45-55; over 65 years old), using the statistical treatment R (R Core Team, 2021). The general results confirm the hypothesis raised, since the percentage of application of the NA rule made up 17.3% of the data, being intermediate to the results of Afro-Brazilian Portuguese and Salvador's Portuguese. Furthermore, in the rural area of the communities under study, the application of the rule is less favored, a behavior similar to what is observed in BP in general. Among the linguistic variables, the syntagmatic configuration of the NP and the number of inflectable NP constituents stand out, revealing that the presence of an item with a semantic plural value favors agreement. The result of phonic salience differs from most studies on NA. As for the phonic salience variable, the result obtained does not favor the application of the rule, contrary to some works on the subject (LOPES, 2001; ANDRADE, 2006). Statistical analysis of social variables indicates that men are more in agreement than women, and that the school does not play such a strong role as an agent of diffusion of the prestige norm. The age group results indicate an acquisition trend with younger speakers applying the agreement rule more than older ones speakers. To explain the correlation between the typological characteristics of Portuguese and the Bantu languages, I adopt the perspective of the evolution of languages from the competition and selection of features (MUFWENE, 2002, 2008) and the null theory (ABOH, 2015; ABOH and DEGRAFF, 2017). I argue that, during the period of colonization of Brazil, there would have been a process of grammatical reanalysis in the acquisition of Portuguese as a second language in which the prefixes of the Bantu languages were reinterpreted following the initial maintenance of the grammar of agglutination of the functional elements observed in the Bantu languages, which would justify marking the plural only in the first element of the NP.

**Keywords:** nominal agreement; portuguese from the inland cities of Bahia; sociolinguistics; language contact.

## LISTA DE ABREVIATURAS

ADJ – adjetivo

ADV – advérbio

CL - classe

CN – concordância de número

DLPs – dados linguísticos primários

DET - determinante

FEM – feminino

GU – gramática univeral

LA – língua alvo

L1 – primeira língua

L2 – segunda língua

MOD - modificador

NUM – numeral

P4 – primeira pessoa do plural

PB – português brasileiro

PE – português europeu

PL - plural

POSS – possessivo

PREF - prefixo

QUANT – quantificador

SING - singular

SN – sintagma nominal

SUBS - substantivo

TLI – transmissão linguística irregular

## LISTA DE FIGURAS

Figura 1 - Localização das comunidades de Poções e Santo Antônio de Jesus.....	99
Figura 2 - Relações entre as medidas de <i>odds</i> , <i>logodds</i> e probabilidade.....	116
Figura 3 - Representação de banco de traços linguísticos.....	175
Figura 4 - A CN no <i>continuum</i> sociolinguístico no português da Bahia.....	180

## LISTA DE GRÁFICOS

Gráfico 1 - Comparação dos resultados de Helvécia e dos Tongas quanto à variável posição linear do item em relação ao núcleo e função (BAXTER, 2009, p. 278).....	49
Gráfico 2 - Resultado geral da aplicação da regra de CN no português popular do interior da Bahia.....	119
Gráfico 3 - Proporções da variável número de constituintes do SN em relação à aplicação da regra de CN.....	131
Gráfico 4 - Frequência de marcação de plural nos itens em primeira posição no SN.....	144
Gráfico 5 - Proporções da variável <i>localidade</i> em relação à aplicação da regra de CN.....	153
Gráfico 6 - Aplicação da regra de CN nas zonas rurais e urbanas de Poções e de Santo Antônio de Jesus.....	178
Gráfico 7 - Distribuição da proporção de aplicação da regra de CN no português popular da Bahia quanto à faixa etária.....	183

## LISTA DE QUADROS

Quadro 1 - Níveis de saliência fônica estabelecidos por Scherre (1988).....	32
Quadro 2 - Resultados da variável marcas precedentes nos dialetos de Helvécia e dos Tongas.....	50
Quadro 3 - Distribuição demográfica da população no Brasil colonial.....	81
Quadro 4 - Distribuição da sociedade brasileira considerando o <i>continuum</i> rural < rurbano < urbano proposta por Lucchesi (2015).....	96
Quadro 5 - Fatores da variável <i>configuração sintagmática do SN</i> .....	101
Quadro 6 - Fatores da variável <i>realização do núcleo</i> .....	104
Quadro 7 - Fatores da variável <i>número de constituintes do SN</i> .....	104
Quadro 8 - Fatores da variável <i>número de constituintes flexionáveis do SN</i> .....	105
Quadro 9 - Fatores da variável <i>saliência fônica</i> .....	105
Quadro 10 - Fatores da variável <i>classe gramatical do constituinte</i> .....	106
Quadro 11 - Fatores da variável <i>posição linear do constituinte marcado</i> .....	106
Quadro 12 - Níveis de saliência fônica (SCHERRE, 1988) e as ocorrências no português popular do interior da Bahia.....	138
Quadro 13 - Prefixos nominais do proto-bantu, segundo Meinhof (1910).....	161
Quadro 14 - Classes, prefixos nominais e significados nas línguas bantu.....	162
Quadro 15 - Numerais cardinais de 1 a 10, no kimbundo, e SNs correspondentes....	168
Quadro 16 - Características das línguas bantu e do português.....	170

## LISTA DE TABELAS

Tabela 1 - Estimativas do modelo 1 (de regressão logística, modelo linear generalizado de efeitos mistos) para a CN no português popular da Bahia.....	121
Tabela 2 - Estimativas do modelo 2 (de regressão logística, modelo linear generalizado de efeitos mistos) para a CN no português popular da Bahia.....	122
Tabela 3 - Rodada 1. Frequência e proporção da variável <i>configuração sintagmática do SN</i> sobre a aplicação da CN no português popular do interior da Bahia.....	124
Tabela 4 - Rodada 2. Frequências e proporções da variável <i>configuração sintagmática do SN</i> sobre a aplicação da CN.....	126
Tabela 5 - Frequências e proporções da variável <i>número de constituintes do SN</i> .....	131
Tabela 6 - Frequências e proporções da variável <i>número de constituintes flexionáveis do SN</i> sobre a aplicação da CN.....	133
Tabela 7 - Frequências e proporções da variável <i>realização do núcleo</i> sobre a aplicação da CN.....	135
Tabela 8 - Frequências e proporções da variável <i>saliência fônica</i> sobre a aplicação da CN.....	137
Tabela 9 - Resultado descritivo da <i>classe gramatical do constituinte</i> na primeira posição do SN.....	140
Tabela 10 - Resultado descritivo da <i>classe gramatical do constituinte</i> na primeira e na segunda posição do SN.....	142
Tabela 11 - Resultado descritivo da <i>posição do constituinte marcado</i> .....	143
Tabela 12 - Frequências e proporções da variável <i>sexo</i> sobre a aplicação da CN.....	146

Tabela 13 - Frequências e proporções da variável <i>faixa etária</i> sobre a aplicação da CN.....	147
Tabela 14 - Frequências e proporções da variável <i>escolaridade</i> sobre a aplicação da CN.....	148
Tabela 15 - Cruzamento da <i>faixa etária</i> com <i>escolaridade</i> .....	150
Tabela 16 - Frequências e proporções da variável <i>estadia fora da comunidade</i> sobre a aplicação da CN.....	150
Tabela 17 - Frequências e proporções da variável <i>localidade</i> sobre a aplicação da CN.....	155

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>18</b>
<b>2. O QUE MOVE A PESQUISA: O OBJETO DE ESTUDO.....</b>	<b>23</b>
2.1 A CONCORDÂNCIA NOMINAL.....	23
2.2 A CONCORDÂNCIA DE NÚMERO NO PORTUGUÊS.....	28
<b>2.2.1 O português urbano do Rio de Janeiro: fixando variáveis / o parâmetro dos estudos de concordância nominal.....</b>	<b>30</b>
<b>2.2.2 A concordância nominal no português do Estado da Bahia.....</b>	<b>38</b>
2.2.2.1 O português de Salvador.....	39
2.2.2.2 O português rural afro-brasileiro baiano.....	45
2.3 SÍNTESE.....	51
<b>3. O CONTATO LINGÜÍSTICO NA FORMAÇÃO DO PORTUGUÊS BRASILEIRO.....</b>	<b>53</b>
3.1 TEORIAS SOBRE A AQUISIÇÃO DE LÍNGUA EM SITUAÇÃO DE CONTATO.....	53
<b>3.1.1 Bioprograma da Linguagem.....</b>	<b>56</b>
<b>3.1.2 Relexificação.....</b>	<b>59</b>
<b>3.1.3 Transmissão Linguística Irregular.....</b>	<b>62</b>
<b>3.1.4 Competição e Seleção – Ecolinguística.....</b>	<b>64</b>
<b>3.1.6 Teoria Nula.....</b>	<b>68</b>
<b>3.1.6 Uma outra proposição aquisicional.....</b>	<b>72</b>
3.2 PRINCIPAIS ASPECTOS SÓCIO-HISTÓRICOS DA FORMAÇÃO DO BRASIL.....	75
<b>3.2.1 Os povos e as línguas que aqui habitavam e que aqui chegaram.....</b>	<b>76</b>
3.2.1.1 As nações indígenas e suas línguas.....	77
3.2.1.2 Os povos africanos e suas línguas.....	82
<b>3.2.2 Os efeitos do contato na urbanização do Brasil e a dicotomia <i>rural-urbano</i>.....</b>	<b>87</b>
3.3 SÍNTESE.....	89



<b>4. ASPECTOS TEÓRICO-METODOLÓGICOS.....</b>	<b>91</b>
4.1 A SOCIOLINGUÍSTICA.....	91
4.2 A ORIGEM DOS DADOS.....	95
<b>4.2.1 As comunidades: Poções e Santo Antônio de Jesus.....</b>	<b>97</b>
4.3 ENVELOPE DA VARIAÇÃO.....	99
<b>4.3.1 A variável dependente .....</b>	<b>100</b>
<b>4.3.2 As variáveis estruturais.....</b>	<b>101</b>
<b>4.3.3 As variáveis sociais.....</b>	<b>107</b>
4.4 O SUPORTE ESTATÍSTICO COMPUTACIONAL E O TRATAMENTO DOS DADOS.....	108
4.5 SÍNTESE.....	117
<b>5. ANÁLISE DOS DADOS.....</b>	<b>118</b>
5.1 RESULTADOS GERAIS.....	118
5.2 O ENCAIXAMENTO LINGUÍSTICO DA CN.....	123
<b>5.2.1 Variável <i>configuração sintagmática do SN</i>.....</b>	<b>123</b>
5.2.1.1 O quantificador tudo.....	128
<b>5.2.2 Variável <i>número de constituintes do SN</i>.....</b>	<b>130</b>
<b>5.2.3 Variável <i>número de constituintes flexionáveis do SN</i>.....</b>	<b>133</b>
<b>5.2.4 Variável <i>realização do núcleo</i>.....</b>	<b>135</b>
<b>5.2.5 Variável <i>saliência fônica</i>.....</b>	<b>136</b>
<b>5.2.6 Variáveis <i>classe gramatical do constituinte marcado e posição da marca de plural</i>.....</b>	<b>140</b>
5.3 O ENCAIXAMENTO SOCIAL DA CN.....	144
<b>5.3.1 Variável <i>sexo</i>.....</b>	<b>145</b>
<b>5.3.2 Variável <i>faixa etária</i>.....</b>	<b>147</b>
<b>5.3.3 Variável <i>escolaridade</i>.....</b>	<b>148</b>
<b>5.3.4 Variável <i>estadia fora da comunidade</i>.....</b>	<b>150</b>
<b>5.3.5 Variável <i>localidade</i>.....</b>	<b>151</b>
5.4 SÍNTESE.....	155
<b>6. O PORTUGUÊS POPULAR DO INTERIOR DA BAHIA E AS LÍNGUAS BANTU.....</b>	<b>158</b>
6.1 A CONCORDÂNCIA NOMINAL NAS LÍNGUAS BANTU.....	158

6.2 INFERÊNCIAS SOBRE O PAPEL DAS LÍNGUAS BANTU NA CONCORDÂNCIA NOMINAL DO PORTUGUÊS RURAL DA BAHIA.....	170
6.3 SÍNTESE.....	176
<b>7. CONSIDERAÇÕES FINAIS.....</b>	<b>178</b>
<b>REFERÊNCIAS.....</b>	<b>185</b>
<b>APÊNDICE.....</b>	<b>197</b>

## 1. INTRODUÇÃO

Partindo do princípio da heterogeneidade linguística do português brasileiro (PB) e de suas características sociais, disponho-me a apresentar em minhas primeiras palavras os pressupostos que fundamentam esta pesquisa de doutorado e que, naturalmente, percorrerão desde a descrição do objeto de estudo e os procedimentos teórico-metodológicos até as ações encadeadas pela pesquisa.

Notadamente, o processo de formação sócio-histórica do PB, no que diz respeito ao contato entre línguas, ocorrido no período de sua formação, tem sido objeto de discussão permanente dos estudiosos da área da Linguística. Adolfo Coelho, em 1880<sup>1</sup>, foi o primeiro a pressupor a influência das línguas africanas na formação do PB devido à existência de características semelhantes em línguas crioulas afro-portuguesas. No entanto, essa hipótese se fortaleceu a partir do estudo de Guy (1981), no qual o autor atrela a formação do PB a um processo de crioulização e rápida descrioulização, em decorrência de fatores sócio-históricos.

Diversos fenômenos linguísticos no PB foram objeto de estudo considerando esse contexto, como por exemplo: (i) a variação de concordância de número e de gênero no sintagma nominal (*os meninos/os menino; cada uma tem uma natureza/cada um tem um natureza*, respectivamente) (Guy, 1981; Lopes, 2001; Baxter, 2009; Lucchesi, 2000, 2009; entre outros); (ii) a variação de concordância entre verbo e sujeito (*Eles ganham/Eles ganha*) (Almeida, 2006; Lucchesi, 2009; Brandão e Vieira, 2012; entre outros); (iii) o uso variável da concordância nominal em predicativos do sujeito e em estruturas passivas (*as coisas tão caras/as coisa tá cara; A casa de farinha tá boa/A casa de farinha tá bom*) (Antonino, 2007, 2012; Lucchesi, 2009; entre outros); (iv) a alternância dativa - dupla possibilidade de realização dos verbos transitivos diretos e indiretos - (*Gente de tempo é assim, se eu dou o senhor marcriação/Gente antiga é assim, se eu faço malcriação para o senhor*) (Lucchesi e Mello, 2009; entre outros).

Os resultados desses trabalhos diagnosticam a necessidade de se considerar o quadro de multilinguismo existente no Brasil nos três primeiros séculos de sua

---

<sup>1</sup> Coelho (1880 *apud* Tarallo, 1993).

história, principalmente, quando houve um intenso tráfico de pessoas, e milhares de africanos foram trazidos para o Brasil como escravizados (LUCCHESI, 2002).

Lucchesi (2000) defende que o movimento de estratificação social da população brasileira contribuiu para a existência de duas grandes normas linguísticas no país: a norma popular, concentrada nos falares interioranos, nas zonas rurais e nas periferias; e a norma culta, mais frequente entre os falantes das classes média e alta dos centros urbanos.

Tendo em vista essa configuração demográfica e, para testar as hipóteses de Bortoni-Ricardo (1998, 2005) e Lucchesi (2015) a respeito do *continuum* linguístico, descrevo e analiso, neste trabalho, o português do interior da Bahia, nos municípios de Poções e Santo Antônio de Jesus. A hipótese do *continuum* permite inferir que essas comunidades seriam intermediárias no que tange à aquisição da norma culta do PB, em que de um lado está o português rural afro-brasileiro baiano (ANDRADE, 2003, 2006; BAXTER, 2009) e do outro, o português popular urbano de Salvador (LOPES, 2001).

Para verificar até que ponto o português popular do interior da Bahia possui características resultantes de contato linguístico do português com as línguas africanas, principalmente (BAXTER, 1992; FIGUEIREDO, 2004; ANTONINO, 2007, LUCCHESI e BAXTER, 2009), tomo como objeto de análise a variação do fenômeno morfossintático da concordância nominal de número (CN) no âmbito do sintagma nominal (SN). Os exemplos em (1) ilustram o fenômeno em estudo: a) apresenta um SN com concordância, pois a marcação de plural está realizada morfossintaticamente nos dois constituintes, e b) apresenta um SN sem concordância, pois a marca de plural está apenas no primeiro constituinte.

- (1) a. *uns policiais* veio trazê ela aqui  
 b. *os pai* perdeu o controle *dos filho* [ANTONM1]

A regra variável de CN é um dos fenômenos mais estudados no português em todas as suas variedades, que colocam em evidência a distinção entre o PB e o PE (NARO e SCHERRE, 1993; 2007). No PE, há uma presença maior, quase categórica, da marcação de plural em todos os constituintes nominais flexionáveis; em contrapartida, no PB, principalmente na variedade popular, as marcas de plural

tendem a ser mantidas apenas no primeiro elemento do sintagma (GUY, 1981; SCHERRE, 1988; LOPES, 2001; ANDRADE, 2003, 2006; BAXTER, 2009).

Estudar a CN no PB é atentar-se, sobretudo, para os fatores extralinguísticos que confirmam a existência de uma polarização linguística e social do PB (LUCCHESI, 2000, 2001, 2009, 2015). De acordo com o autor, o fenômeno afeta os dois extremos das normas linguísticas, porém há uma intensidade maior na variedade popular, na qual é mais frequente o apagamento da marca de número nos nomes núcleos do SN. Nesse sentido, a presente pesquisa também envolve a problemática de responder se os padrões da regra de concordância no português do interior da Bahia seguem os pressupostos de um *continuum*, com maior aplicação da regra do que no português afro-brasileiro baiano e menor do que no português urbano de Salvador

Os resultados indicam que o português popular do interior da Bahia é uma variedade intermediária, isto é, rurbana (BORTONI-RICARDO, 1998, 2005; LUCCHESI, 2015), pois o percentual geral de aplicação da regra de CN foi de 17,3%: maior do que no português afro-brasileiro baiano, 8% (ANDRADE, 2006) e menor do que no português de Salvador, 34% (LOPES, 2001).

Ademais, considero que o tema desta pesquisa é relevante por se tratar de um estudo que poderá, através da análise linguística e extralinguística da CN, verificar se o processo da (não) marcação de plural reflete correlação existente entre as características tipológicas das línguas africanas, crioulas e o português, já que consideramos o comportamento da regra variável de CN no PB como resultante do processo contato linguístico com línguas africanas, principalmente. Assim, busco verificar se os substratos africanos explicariam a CN no português do interior da Bahia, e em que medida os resultados da comparação de outros estudos sobre o PB podem elucidar as peculiaridades do português adquirido em situação de contato linguístico.

Parto da hipótese central de que o português do interior da Bahia apresenta um quadro de mais ausência do que presença de aplicação da regra de concordância por se tratar de uma variedade linguística popular influenciada por fatores sócio-históricos característicos de comunidades de fala interioranas, marcadas pelo contato de indígenas e negros entre si e entre a população branca e mestiça, ocasionando no distanciamento da norma culta do PB (BORTONI-RICARDO, 2005).

O aporte teórico-metodológico utilizado na investigação da CN neste trabalho é o da Sociolinguística Variacionista (LABOV, 2008 [1972]), a qual estabelece a relação entre o comportamento linguístico de uma determinada comunidade de fala com os

fatores sociais e com os fatores estruturais da língua. Considero relevante também apresentar as proposições das teorias de contato, das precursoras às mais recentes.

No Capítulo 2 deste trabalho, descrevo o objeto de estudo no que diz respeito à categoria funcional nas línguas naturais e sua natureza variável como fenômeno linguístico no português. Apresento os resultados relevantes de algumas pesquisas já realizadas sobre tema no PB (SCHERRE, 1988; LOPES, 2001; ANDRADE, 2003, 2006; BAXTER 2009), as quais servem de modelo para os procedimentos analíticos desta tese. No Capítulo 3, faço uma breve explanação sobre a história do povoamento do Brasil, descrevendo as abordagens excepcionalistas e não-excepcionalistas das teorias sobre a aquisição de língua em situação de contato. Ressalto que o presente trabalho focaliza a história social que levou ao contato linguístico e à formação do PB a partir de modelos teóricos não-excepcionalistas, na medida em que não considero o uso da CN no PB popular como simplificação morfossintática.

Em seguida, no Capítulo 4, apresento o suporte teórico-metodológico da pesquisa, a Sociolinguística Variacionista, faço a revisão dos conceitos básicos referentes a essa teoria, bem como dos procedimentos considerados necessários para uma investigação sociolinguística. Ainda nesse Capítulo, comento os pormenores da origem e do tratamento dos dados e os fatores linguísticos e sociais utilizados para a análise da CN, descrevendo também o suporte computacional R, utilizado para análise estatística dos dados.

No Capítulo 5, faço a exposição dos resultados e a análise dos dados, comentando a confirmação ou não das hipóteses levantadas para fatores linguísticos e sociais da CN, além de cotejar os resultados obtidos neste trabalho com outros já realizados sobre o tema, sobretudo no Estado da Bahia (LOPES, 2001; ANDRADE, 2003, 2006; BAXTER 2009). A partir dos resultados apresentados no Capítulo 5, levanto os seguintes questionamentos:

1) A variação na aplicação da regra de concordância no PB pode estar relacionada ao contato linguístico ocorrido no período de colonização do Brasil?

2) Teriam as línguas de substrato, principalmente as da família bantu, que foram as línguas mais faladas em território baiano (LUCCHESI, 2002; GUY, 2005; PESSOA DE CASTRO, 2006; PETTER, 2006), interferido na marcação de plural apenas no primeiro elemento do SN, um fenômeno observado no PB?

As respostas para esses questionamentos compõem o Capítulo 6, no qual testo a hipótese de Guy (2005) sobre a interferência das línguas africanas do grupo bantu

na marcação de plural do PB. Além disso, apresento a proposta desta tese, atrelando aspectos linguísticos e sócio-históricos das línguas em análise, considerando a hipótese do contato na perspectiva da evolução das línguas a partir da competição e da seleção de traços (MUFWENE, 2002, 2008) e a teoria nula (ABOH, 2015; ABOH e DEGRAFF, 2017).

Tendo em vista que o PB é uma língua transatlântica (ALENCASTRO, 2000) resultante de um contato linguístico intenso e extenso sob condições sócio-históricas de colonização portuguesa, genocídio de povos indígenas e escravização de africanos, presumo que estiveram disponíveis para os aprendentes de português como segunda língua (L2) conjuntos de traços linguísticos tipologicamente diferentes que entraram em competição e em uma posterior seleção direcionada, principalmente, por fatores sociais. Proponho, como tese, que características aglutinantes da gramática das línguas bantu, sobretudo do kimbundo, foram incorporadas ao português, mantendo o plural prefixado ao nome, ou seja, na posição inicial do SN.

Após as discussões do Capítulo 6, apresento as considerações finais da pesquisa, ressaltando seus principais objetivos e os resultados alcançados. Destaco, neste ponto do trabalho, a verificação do *status* da CN no *continuum* linguístico no Estado da Bahia e evidencio que a temática deste estudo não se esgota aqui.

## 2. O QUE MOVE A PESQUISA: O OBJETO DE ESTUDO

Com o objetivo de apresentar a definição e o estado da arte da CN, neste capítulo, descrevo o objeto de estudo desta pesquisa enquanto mecanismo funcional, considerando a realização da marcação de número no PB e exponho os resultados de algumas pesquisas variacionistas, as pioneiras e as mais recentes, sobre a CN no português. Por fim, comento as matrizes metodológicas e os resultados dessas pesquisas que são relevantes para este trabalho.

### 2.1 A CONCORDÂNCIA DE NÚMERO

A concordância é um fenômeno linguístico em que duas ou mais palavras concordam em relação a algum traço gramatical com o mesmo valor de relevância entre elas (RADFORD, 1997, p. 492)<sup>2</sup>. Conforme Trask (2004), a concordância não é um fenômeno universal, isto é, não é comum a todas as línguas naturais, apesar de estar presente em muitas delas. Ele elenca as línguas em três grupos distintos considerando a maneira como indicam pluralidade: (i) línguas que possuem alto índice de concordância, como o suaíli, o russo, o latim, o alemão; (ii) línguas em que a concordância pode ser realizada com a marcação de plural em apenas um elemento do sintagma, como o francês, o português, o espanhol e o inglês; (iii) e línguas que não possuem marcas flexionais de concordância, como o chinês.

Nas línguas em que há concordância, é possível verificá-la observando a relação de traços gramaticais entre os itens constituintes do SN, e entre sujeito e verbo ou objeto e verbo. As características gramaticais envolvidas na relação de concordância entre as palavras de uma sentença implicam a operação morfológica de flexão que é a possibilidade de modificação de determinadas classes de palavras para expressar diferentes categorias gramaticais, como gênero, número, pessoa e caso.

---

<sup>2</sup> “Agreement: two words (or expressions) are said to agree in respect of some grammatical feature(s) if they have the same value for the relevant feature(s): so, in a sentence such as He smokes, the verb *smokes* is said to agree with its subject *he* because both are third person singular expressions.” (RADFORD, 1997, p. 492)



Verifica-se que a concordância pode ser nominal, na qual as palavras podem concordar em gênero, número e caso no âmbito do SN, como ocorre no exemplo (1); em número e pessoa entre o SN sujeito e o verbo, como em (2); e com o objeto (3).

- (1) a. **A**            menina            bonita  
       DET **FEM** SIN    SUBS **FEM** SING    ADJ **FEM** SING
- b. **As**            meninas            bonitas  
       DET **FEM** **PL**    SUBS **FEM** **PL**    ADJ **FEM** **PL**
- c. **Bonae**            discipulae  
       ADJ **PL-Boas**    SUBS **PL-discipulas**

- (2) **Nós**    ganha-**mos**    o prêmio.  
       **P4**        VERBO- **P4**

- (3) a. 'o-wolw-ii  
       /3Sg-speak-PAST/  
       'he spoke'  
       (Ele falou)
- b. cf. 'o-wolw-**an**-ii moodibbo  
       /3Sg-speak-**Ext**-PAST teacher/  
       'he spoke **to** the teacher'  
       (Ele falou com o professor)

(PAWLAK, 2012, p. 160)

Como se vê, no exemplo em (1a), o SN apresenta um traço de gênero [+feminino] representado pelo morfema *-a* em todos os constituintes. Em (1b), o morfema *-s* indica plural em todos os itens do SN. Em (1c), há um exemplo de concordância do caso nominativo no latim, indicado pela declinação *-ae*. Em (2), há um exemplo de concordância verbo-sujeito, em que a desinência verbal *-mos* concorda em número-pessoa com o pronome nós (primeira pessoa do plural). Os exemplos elencados em (3), mostram a concordância verbo-objeto na língua africana Fulfude, em que o morfema *an* corresponde a uma extensão verbal para indicar que o seu complemento é um objeto indireto (Pawlak, 2012).

Quanto às línguas que demonstram um sistema de concordância maior, caracterizam-se por apresentar, obrigatoriamente, em todos os constituintes da sentença algum tipo de marcação morfológica que indique número, gênero, pessoa e caso (Corbett, 2000; Trask 2004), como o alemão (4).

- (4) *Die        hübschen        Mädchen        leben        hier.*

DET PL-As    ADJ PL-bonitas    SUBS PL- meninas    VERBO PL-vivem    ADV-aqui.

As línguas caracterizadas com menos concordância são as que apresentam um sistema reduzido de marcação morfológica das categorias gramaticais, como o francês (5), em que os morfemas de plural são marcados grafematicamente, porém não soam na língua falada. Outra língua em que a marcação de plural não é necessária em todos os constituintes do sintagma é o inglês (6), em que apenas o núcleo do SN recebe a marcação de plural, enquanto os determinantes e os modificadores permanecem no singular.

(5)    *Les*                    *vieux*                    *livres*  
 DET PL-Os            ADJ PL-antigos            SUBS PL-livros

(CASTILHO, 2010, p.401)

(6)    *The*                    *five*                    *new*                    *houses*  
 DET SG-A            NUM-cinco            ADJ PL- novas            SBUS PL-casas

Nas línguas que não possuem categoria morfológica para marcação de concordância, como afirma Ao (2013), a noção de singular ou plural é expressa pela ordem dos constituintes na sentença ou pela presença de partículas à parte dos vocábulos, como o chinês (7).

(7)    neoihaaizi            **dou**                    coeng-dak            han                    hou teng  
       女孩子            **都**                    唱 -得很            好                    聽。  
       SUBS -Rapariga    **quant.**            VERBO-cantar    ADV-muito            ADJ-bem  
       (As raparigas cantam muito bem)

(AO, 2013, p.31)

Nas línguas românicas, a CN diz respeito à combinação das categorias gramaticais de gênero e número com aspectos semântico-lexicais, como mostra o exemplo (8) para o português, o espanhol e o francês, respectivamente.

(8)    a. As pequenas meninas  
       b. *Las niñas*  
       c. *Les petites filles*

No caso específico da concordância de número, a qual é determinada por flexão morfológica indicativa de pluralidade, a maneira como as línguas naturais

podem apresentar pluralidade nos nomes é distinta. Corbett (2000) elenca algumas maneiras: (i) por prefixação, em línguas como *aindilyakwa*, do norte da Austrália; (ii) por sufixação, como no inglês e no português; (iii) por alterações na raiz da palavra, como ocorre no *sudanês* e no *dinka*; (iv) pela entonação, como algumas línguas africanas, *gworok*, *seme* e *beria*; (v) reduplicação completa da raiz, como em *indonésio*; (vi) pela combinação de dois ou mais métodos, como por exemplo, as línguas *misantla* e *totonac*, do México, que marcam o plural dos nomes com os processos de prefixação e de sufixação. Há, segundo o autor, ainda línguas cujo plural é marcado em todos os nomes contáveis, em outras a marcação de plural limita-se a nomes com traço [+animado] ou com traço [+humano].

Corbett (2000), porém, ressalta que os processos de pluralização podem ser resumidos de dois modos: o primeiro diz respeito à mudança morfológica no nome, como ocorre em português (*casa/casas*), com a inserção de um morfema indicativo de plural; no outro modo, a pluralidade é marcada com um morfema adjunto a alguma palavra dentro do SN, como ocorre em (9), exemplo de SN da língua havaiana.

- (9) 'elua            a'u        **mau**        i'a  
           two            my        **PL**        fish  
           NUM-dois    POSS-meu    **quant.**    SUBS-peixe  
                   'Meus dois peixes'

(ELBERT E PUKUI, 1979 *apud* CORBETT, 2000)

No português, a marcação de número tem dois valores: singular e plural. A forma singular dos nomes não é marcada morfológicamente, enquanto a forma plural ocorre, pela regra geral, através do processo de sufixação com a inserção do morfema – s. Em (10), podemos ver a aplicação da regra de concordância de número através da marcação morfológica com morfema -s, sendo essa expressão mórfica estendida nos itens flexionáveis do SN. Cabe salientar que o morfema -s possui aloformes como -es e -is, por exemplo, os quais são condicionados foneticamente.

Neste trabalho, analiso a variação da CN no SN em amostras de fala do português de Poções e Santo Antônio de Jesus, levando em consideração os aspectos morfofonológicos e morfossintáticos que envolvem esse processo. No PB, a CN é um fenômeno variável – ora verifica-se a aplicação da regra, como os exemplos em (10) em que todos os elementos do SN estão marcados para plural, ora apenas o primeiro termo do SN detém a marca de número, como em (11).

- (10) a. ah, tinha *os musgos* vinha tudo praí [ISAULF3]  
 b. nós estava sozinha den'de casa, nós quatro, *as irmãs mais novas* [ROSF1]
- (11) a. *os vizinho* foi tudo lá pra casa [ANTONM3]  
 b. *os pai* perdeu o controle dos filho [ANTONM1]

Estudos sobre a CN no PB têm revelado que a (não) aplicação da regra de concordância envolve não só processos morfossintáticos, mas também processos fonéticos-fonológicos. Quanto aos processos fonéticos-fonológicos, sabe-se que, no português, pode haver o contraste de material fônico quando comparadas as formas singular e plural de um dado item (LEMLE e NARO, 1976). Esse contraste é denominado de saliência fônica e, segundo Chaves (2014), é um fator de natureza morfofonológica apontado na literatura como uma das variáveis mais significativas no que diz respeito à alternância na marcação de número.

Quanto ao processo morfossintático, a CN envolve a análise do morfema de plural nos vocábulos a partir da perspectiva de que esse constituinte é preferencialmente fixado nos itens que se localizam à esquerda do núcleo do SN (BAXTER, 2009). Outro ponto é a análise da inserção do morfema de plural no sintagma que ocorre primeiro na estrutura do determinante e que pode definir o material lexical do SN, ou seja, o SN como um todo seria um complemento do determinante (ABNEY, 1987).

Considerando esses aspectos: fonético-fonológico e morfossintáticos, os trabalhos sobre CN têm investigado, de forma geral, as seguintes variáveis em relação à regra de concordância: saliência fônica imposta pela aplicação da regra, classe gramatical dos constituintes do SN, posição do constituinte no SN na sentença, (não) existência de marcas de número precedentes ao constituinte, configuração da estrutura do SN, dentre outros. Essas variáveis são detalhadas na resenha dos trabalhos selecionados como relevantes para este trabalho, bem como para a fixação das variáveis utilizadas para a descrição dos dados deste trabalho.

Dentre diversos estudos já realizados<sup>3</sup>, selecionei os que contribuem para o aprimoramento desta pesquisa na perspectiva teórico-metodológica, além de tratarem do tema em comunidades de fala com características semelhantes aos municípios

---

<sup>3</sup> Por exemplo: Scherre, 1978, 1988; Guy, 1981, 2000; Naro & Scherre, 1991; Fernandes, 1996; Salomão, 2010; Brandão & Vieira, 2012; Yacovenco et. al. 2012; Silva & Scherre, 2013; Mendes, 2014; Oushiro, 2015; entre outros.

baianos de Poções e de Santo Antônio de Jesus: Scherre (1988), Lopes (2001), Andrade (2003, 2006), Baxter (2009), Figueiredo (2010) e Brandão (2015).

A seleção dos trabalhos visa contribuir para o alcance dos seguintes objetivos: i) verificar quantitativamente a existência de um *continuum* entre o português rural afro-brasileiro (Andrade, 2003, 2006; Baxter, 2009) e o português popular de Salvador (Lopes, 2001), em que a aplicação da regra de concordância é menos frequente naquele do que nesse; ii) comparar o comportamento morfossintático da CN no português do interior da Bahia com outras variedades do PB; iii) buscar evidências que permitam inferir acerca das características sociolinguísticas da CN.

Antes, porém, de tentar responder às questões que levaram aos objetivos propostos no parágrafo anterior, passo à descrição da CN no português, mostrando os principais resultados já alcançados em pesquisas sociolinguísticas.

## 2.2 A CONCORDÂNCIA DE NÚMERO NO PORTUGUÊS

A CN entre elementos do SN é um fenômeno linguístico amplamente estudado no que diz respeito à comparação do PB com o PE e se caracteriza como um fator que distingue essas duas línguas (NARO, 1981; SCHERRE, 1988; NARO & SCHERRE, 1993, 2000, 2007; BRANDÃO, 2013, 2015). No PE, a regra de aplicação da CN é aplicada na forma canônica quase que de forma categórica, apresentando um percentual muito baixo de variação mesmo entre falantes que possuem pouca ou nenhuma escolaridade e indica um quadro de estabilidade linguística (NARO E SCHERRE, 2007; BRANDÃO, 2015).

Questionando-se a respeito do forte contraste entre o PE e as variedades não europeias do português quanto ao uso da CN, Brandão (2015), em um estudo comparativo, constatou que, para o PE, a regra de CN é aplicada conforme a regra padrão por todos os grupos de falantes, sendo encontrados apenas 8 casos de ausência da marca de plural dentre 6.952 ocorrências. A autora destaca que esse fato se caracteriza no que Labov (2003) chamou de lapso, visto que a regra de aplicação é quase categórica. Como forma de evidência, Brandão (2015) explicita isso com dados, como os exemplos (12) e (13), em que o falante reorganizou o discurso imediatamente, marcando o plural.

(12) a maioria **d[as pessoa] das mulheres** era tra/ era em casa era a bordar

(13) **[outros conhecimento] outra maneira** de ver as pessoas

(BRANDÃO, 2015, p. 39)

Em contrapartida, o PB apresenta variação quanto à aplicação da CN, que envolve, principalmente, o processo de apagamento da marca de plural, /s/, nos nomes na fala de pessoas menos escolarizadas. A variação observada no PB, no que diz respeito à variação na CN, poderia apontar para um quadro de variação estável ou de mudança em curso em direção à perda de concordância, visto ter sido o PE, uma variedade em que a regra categórica é a aplicação da concordância, o modelo para a formação do PB. No entanto, a maioria dos estudos já realizados, contrariamente, tem mostrado um quadro de mudança em curso no que tange à aquisição da regra de concordância como consequência da ação de fatores sociais, como a escolarização (SCHERRE, 1988; LOPES, 2001; ANDRADE, 2003, 2006; VIEIRA & BRANDÃO, 2014; LOPES, 2014; RIBEIRO, 2018, dentre outros).

O conjunto de regras para a CN apresentado nas gramáticas normativas que circulam no Brasil segue o padrão lusitano e não corresponde à realidade do PB. Esse sistema de regras é, de certa forma, redundante, pois uma vez marcado o plural no primeiro elemento do sintagma, subtende-se que todo o SN tenha o valor de plural, não sendo necessária a marcação em todos os elementos flexionáveis. Isso permite a variação no uso de uma ou de outra forma linguística, como questiona Castilho (2010, p. 401): “Por que tanto {-s} se espalhando ao longo do sintagma nominal, como em *os livros antigos*? [...] Basta dizer *os livro antigo* [...]” para que a informação seja entendida.

Considerando que o objetivo deste trabalho também envolve contribuir para a história do PB, no que diz respeito à (não) aplicação da regra de concordância, é relevante ter em vista os trabalhos já realizados sobre o tema. Dessa forma, o objetivo desta seção é resenhar os estudos variacionistas de Scherre (1988), Lopes (2001), Andrade (2003, 2006) e Baxter (2009), a fim de relacioná-los, sobretudo no Capítulo 5, aos resultados obtidos da análise do português do interior do Estado da Bahia.

### 2.2.1 O português urbano do Rio de Janeiro: fixando variáveis / o parâmetro dos estudos de concordância nominal

Tendo em vista a importância do fenômeno da CN no cenário comparativo entre as variedades do português, faz-se necessário uma abordagem dos trabalhos pioneiros de destaque sobre o tema no PB, como é o caso do trabalho de Scherre (1988) que se tornou uma obra de grande referência no assunto por postular variáveis relevantes que regem o fenômeno da variação da CN no SN. Nesta seção, apresento os principais resultados dessa pesquisa desenvolvida pela referida autora e apresentada como tese de doutorado.

Scherre (1988), adotando a Sociolinguística como abordagem teórica, estudou a variação de plural no SN utilizando dados da Amostra Censo 1980 do PEUL/UFRJ, constituída de 64 falantes do Rio de Janeiro, distribuídos em 48 adultos (15 – 71 anos) e 16 crianças (7 – 14 anos), estratificados por sexo, anos de escolarização (1 a 4 anos; 5 a 8 anos; 9 a 11 anos) e idade. Ela analisou o fenômeno a partir de duas perspectivas: uma mórfica ou atomística e outra sintagmática ou não-atomística; Em ambas as análises, foram consideradas duas variantes: presença *versus* ausência da marca de plural.

A perspectiva mórfica leva em conta cada elemento constituinte do SN como um dado de análise e tem por objetivo analisar a marcação de plural no que tange à morfologia de cada palavra, como no exemplo (14) em que o SN possui quatro elementos e cada um se constitui como um dado.

(14) as pernaØ todaØ marcadaØ

(SCHERRE, p. 62)

A perspectiva sintagmática considera todo o SN como uma só unidade de análise, como mostra o exemplo (15). Nessa perspectiva, o objetivo é analisar a CN na estrutura do SN no que se refere à presença ou não da harmonia formal estabelecida pela relação sintática entre os constituintes do sintagma (SCHERRE, 1988, p. 62)

(15) a. aqueles médicos todinhos [+ concordância]

b. aqueles médico todinho [- concordância]

(SCHERRE, p. 288)

As duas perspectivas objetivam mostrar que algumas variáveis atuam especificamente sobre cada elemento do SN, outras atuam principalmente sobre o SN como um todo e há aquelas que atuam tanto sobre os elementos do SN separadamente quanto sobre o SN inteiro (Scherre, 1988, p. 61).

Saliento que, nesse estudo para ambas as perspectivas – mórfica e sintagmática, Scherre (1988) considerou três variáveis sociais: escolarização, faixa etária e sexo do falante, as quais se mostraram relevantes para a análise do fenômeno. Como os resultados são semelhantes nas duas perspectivas de análise, teço comentários sobre eles em conjunto, após a explanação dos resultados das variáveis linguísticas em ambas as análises.

Com o objetivo principal de descrever e explicar a sistematicidade do uso variável da CN no SN e os fatores que favorecem ou não a aplicação da regra, Scherre (1988) elencou onze variáveis linguísticas para analisar a variação de marcas de plural no SN na perspectiva mórfica. Dentre as variáveis linguísticas, oito se mostraram mais significativas à marcação do plural nos itens do SN: saliência fônica (processos morfofonológicos de formação de plural e tonicidade da sílaba dos itens lexicais singulares); relação entre os elementos não nucleares em função do núcleo e posição dos elementos nucleares no SN; marcas precedentes em função da posição; contexto fonético/fonológico seguinte; função resumitiva do SN; formalidade dos substantivos e adjetivos; grau dos substantivos e adjetivos e animacidade dos substantivos. Apresento, a seguir e de forma breve, o comportamento de cada uma dessas variáveis analisadas por Scherre (1988).

A variável saliência fônica parte do princípio de que as formas mais salientes, e por isto mais perceptíveis, são mais prováveis de serem marcadas do que as menos salientes devido à diferenciação fônica (Lemle e Naro, 1976). Scherre (1988) analisou a saliência fônica sob três dimensões:

1) processos morfofonológicos de formação do plural: para analisar as formas que apresentam maior diferenciação de material fônico na relação singular/plural (ovo/ovos; leitão/leitões; casal/casais) em oposição às que apresentam menor diferenciação (coisa/coisas; carioca/cariocas);



2) tonicidade da sílaba dos itens lexicais singulares: para verificar a relação entre a marcação de plural e a acentuação silábica dos vocábulos, considerando como contexto mais saliente os oxítonos singulares e os monossílabos tônicos (pé; pais, país, leitão, principal);

3) número de sílabas dos itens lexicais singulares: para examinar o número de sílabas dos vocábulos considerando como mais salientes os vocábulos maiores, por conterem mais material fônico (salário, português, descartável).

Scherre (1988, p. 75-77) estabeleceu uma escala de níveis de saliência partindo dos itens mais salientes para os menos salientes, conforme o Quadro 1.

Quadro 1. Níveis de saliência fônica estabelecidos por Scherre (1988).

<b>Níveis de saliência fônica</b>	<b>Exemplos</b>
itens com plural duplo	ovo/ovos
itens terminados em <i>-l</i> com plural terminados em <i>-s</i>	casal/casais
itens terminados em <i>-ão</i> , cujo plural se faz com inserção de <i>-s</i>	mão/mãos
itens terminados em <i>-r</i> , cujo plural se faz através da inserção de <i>-es</i>	cor/cores
itens terminados em <i>-s</i> com plural terminados em <i>-es</i>	freguês/fregueses
itens com plural regular, nos quais é feita somente a inserção de <i>-s</i>	lei/leis

Fonte: SCHERRE, 1988, p. 75-77.

A autora destacou a possibilidade de o princípio da saliência atuar de forma diferente de acordo com as características sociais dos falantes, como a escolarização, porém, de maneira geral, constatou que quanto maior a diferença fonológica entre a forma singular e a forma plural de um item lexical, maior a probabilidade desse item receber marcação de plural, como em (16).

(16) nóvos papeizinhos (Joa10,fp,27a)

(SCHERRE, 1988, p. 75)

Como se vê em (16), as palavras *novos* e *papeizinhos* se caracterizam por apresentar plural duplo e terminar com -l, respectivamente, encaixando-se como vocábulos mais salientes nos níveis de diferenciação fônica estabelecidos.

A variável posição linear do constituinte diz respeito ao local em que o constituinte aparece no SN. Scherre (1988) fez uma análise conjunta dessa variável com as variáveis classe gramatical e marcas precedentes, considerando que, de maneira geral, o mecanismo da concordância, no português, segue a tendência de marcação no constituinte em primeira posição no SN e que os artigos são menos propensos à marca zero, enquanto os substantivos e adjetivos são mais propensos a não serem marcados com o plural. Os resultados de Scherre (1988) mostraram que:

1) os itens que se localizam na primeira posição são mais marcados, como o constituinte *as* no exemplo (17):

(17) as boas ações (Ire17,fp,52a)

(SCHERRE, 1988, p. 144)

2) os artigos são a categoria gramatical mais marcada, como podemos observar também no exemplo (17) em que *as* é um artigo definido;

3) os substantivos têm mais chances de serem marcados na terceira posição do que na segunda, conforme o exemplo (18), em que o substantivo *casas* se encontra na terceira posição;

(18) todas as casas (Hel44,fc,44a)

(SCHERRE, 1988, p. 144)

4) os adjetivos se apresentam mais marcados na segunda do que na terceira posição, como se vê em (19), o adjetivo *lindas* em segunda posição recebendo a marca de plural.

(19) coisas lindas (Dor29 ,fg,44a)

(SCHERRE, 1988, p. 144)

Scherre (1988) diz que não se pode estabelecer um paralelo entre determinante e primeira posição, substantivo e segunda posição, adjetivo e terceira posição, mesmo nos SNs formados por três ou mais constituintes, pois, segundo ela, qualquer classe gramatical que esteja na primeira posição tende a ser mais marcada.

Sobre marcas precedentes, Scherre (1988) mostra que a presença de marcas de plural no item anterior influencia significativamente a marcação nos itens posteriores, indicando a tendência que “marcas conduzem a marcas e zeros conduzem a zeros” (p.382), como mostra o exemplo (20), em que o constituinte *feios* está na última posição do SN e recebe a marca de plural assim como os constituintes anteriores a ele.

(20) os piores nomes feios (Ire17,fp,52a)

(SCHERRE, 1988, p. 144)

A variável contexto fonético/fonológico seguinte indica que as marcas de plural são mais frequentes quando o contexto seguinte é consonantal, como em (21), ou quando ocorre uma pausa.

(21) as contas quase toda (Glo11,fg,48a)

(SCHERRE, 1988, p. 191)

Sobre a função sintática do SN, a autora destacou que os SNs que exercem funções de sujeito, adjunto adnominal ou verbal, predicativo do sujeito, recebem mais marcas de plural. O exemplo (22) evidencia isso, pois os constituintes do SN - *as garotas* - recebem marca de plural e exercem função de sujeito.

(22) as garotas brincam (Glo11,fg,48a)

(SCHERRE, 1988, p. 260)

Quanto às variáveis que englobam traços mórficos, semânticos e estilísticos do substantivo ou do adjetivo, Scherre (1988) destaca que para a formalidade dos substantivos, os itens lexicais formais recebem mais marcação de plural do que os informais. A autora considerou como informais os nomes que são gírias ou expressões populares, e como formais os nomes que não se encaixam nessas características. Em (23), há exemplos elencados pela autora como nomes com traços de informalidade e que tendem a não receber marca de plural.

- (23) a. os paraíba brabo lá... (Jae25,mg,30a);  
 b. tomou uns goró, queimou uns fumo e surrou... (Jov26,mg,32a);  
 c. tá fritando uns troço... (Jov26,mg,32a);  
 d. uns cara machão... (Pac20,mg,32a);

- e. acontece cum os otro... (Cla27,mg,30a);  
 f. fazendo essas miséria todinha... (Ire17,fp,52a).

(SCHERRE, 1988, p. 268)

A autora também aponta que os nomes de grau normal apresentam maior marcação. Os exemplos em (24) apresentam esse contraste.

- (24) a. tem umas garotinha lá... (Val24,fg,15a);  
 b. são meus amigão... (Seb09,mg,39a);  
 c. tem umas briguinha... (Ari30,fc,43a);  
 d. sai uns quatrocentos bolinhos de... (Ana40,fc,18a);  
 e. e outras escolinha pequena, eles... (Eve43,fc,42a);  
 f. dois dentão enorme que... (Mal16,fp,56a).

(SCHERRE, 1988, p. 273)

Em relação à animacidade, Scherre (1988) destaca que os nomes com traços [+humano] são mais marcados com o plural em oposição aos [-humano], conforme consta no exemplo (25), a palavra *pais* contém o traço [+humano] e recebe a marca de plural.

- (25) com meus pais eu já viajei... (Val24,fg,15a)

(SCHERRE, 1988, p. 271)

Partindo para a análise sintagmática dos dados, Scherre (1988) analisou as seguintes variáveis estruturais: a configuração sintagmática, a pluralidade do contexto, a função textual do SN, a localização do SN em relação ao verbo e à oração, o *status* informacional do SN, o grau/formalidade do SN, a pluralidade do SN e a animacidade do SN.

A variável pluralidade de contexto considerou o princípio de que os falantes tendam a fazer uma harmonia formal em todos os constituintes não somente do SN, mas também com outros sintagmas, como o verbal. Os resultados mostram que a presença da SNs totalmente marcados ou a presença de pronomes pessoais plurais nas vizinhanças do SN analisado favorece(m) a CN, como ocorre no exemplo (26).

- (26) A maior parte dos meus amigos falam gíria, sabe? Falam mais para gíria. ELES falam (...) Alguma coisa do tipo palavrão também. (Sam01,mp,18a);

(SCHERRE, p. 284)

Sobre a configuração sintagmática do SN, Scherre (1988) considerou a estrutura do SN. Os resultados indicam que a presença de artigo definido na primeira posição do SN e a presença de um substantivo na última posição do SN favorecem a marcação de plural em todos os constituintes do sintagma, conforme exemplo (27).

- (27) a. os meus filhos  
 b. uns bons filhos

(SCHERRE, 1988, p. 308)

Scherre (1988) diz que a quantidade de marcas formais de plural no SN correlaciona-se com a função do tipo de relação que se estabelece entre os elementos dentro do texto. Baseando-se nisso, ela quantificou essa variável, para a qual os resultados apontam que as funções que os SNs menos marcados são aqueles que estabelecem uma referência dentro do texto, remetendo-se explicitamente a elementos já presentes no contexto anterior e os SNs mais marcados são os que não estabelecem referência textual explícita, isto é, as funções que possuem informação não redundante são as que mais favorecem à aplicação da regra de CN nos SNs (p. 313), como podemos ver no exemplo (28).

- (28) a. Meu pai tinha uma...uma preocupação social... Quando a gente foi pra lá, aí casas todas foram arreventadas, essas de sapê foram construídas novas. (Hel44,fc,44a).  
 b. Há uns três anos atrás, nosso barraco... (Joa10,fp,27a);

(SCHERRE, 1988, p. 313; 319; 320)

A variável *status* informacional do SN foi analisada por Scherre (1988) com base na relação entre a maneira de apresentar a informação e a configuração sintagmática dos SNs. Os resultados indicam que os SNs com informações novas recebem mais marcas de plural do que os SNs com informações já mencionadas de algum modo no texto ou fora dele, como se vê em (29), o SN que insere uma nova informação no contexto é totalmente marcado com a CN.

- (29) Eu tava desempregado. (...) Precisava arrumá...uma ocupação... Aí comecei a procurá no jornal... Aí arrumei um... Fui na Uruguaiana. Aí cheguei lá e tinha umas oito ou deyz pessoas na frente. Aí eu esperei a minha vez. (Cla27,mg,32a);

(SCHERRE, 1988, p. 325)

Quanto à localização do SN, Scherre (1988) analisou os SNs no que tange à localização linear do verbo, verificando a questão da topicidade ou da centralidade do SN. Citando Pontes (1986, 1987), Scherre explica que os SNs tópicos ou centrais ocorrem à esquerda da oração ou à esquerda do verbo. Os resultados mostram que os SNs à esquerda da oração ou do verbo receberam maior pluralidade, de acordo com os exemplos em (30).

- (30) a. Ah! O Carnaval é... Carnaval eu gosto. Todos os anos eu vô assisti iscola di samba. (Dor29,fg,44a)  
 b. Todas as ruas, né? Pintaram... (Jos35,fp,59a)

(SCHERRE, 1988, p. 322; 323)

Os resultados da variável grau de formalidade do SN apontam que os SNs formais favorecem mais a CN do que os informais, tal como os resultados da análise mórfica. Qualquer SN que tivesse um item do tipo *cara, outro, besteira, bobagem, grilo, lance, papo, merda, goró, brabão*, etc. ou nomes no diminutivo/aumentativo ou ainda os SNs com núcleo que expressasse função resumitiva, como as construções *um montão de; uma pá de; um bando de; um monte de* (SCHERRE, 1988, p. 331). Para a variável animacidade do SN, Scherre (1988) constata que os SNs com núcleo [+humano] são mais marcados do que os SNs com núcleo [-humano].

Segundo Scherre (1988), os resultados para a variável pluralidade do SN mostram que os SNs que se referem a conceitos, situações, eventos ou entidades que, comumente, são mais de um, favorecem a CN, como podemos verificar nos exemplos (31), os SNs destacados correspondem a conceitos caracterizados em mais de um.

- (31) a. ficam brincando com os nossos direitos. (Hel44,fc,44a);  
 b. nós não damos tanto valor às nossas raízes. (Mgl48,fc,52a).

(SCHERRE, 1988, p. 332)

O estudo de Scherre (1988) é bastante relevante para o tema da CN, pois fixou variáveis significativas para os trabalhos sobre o fenômeno. Apesar disso, em

trabalhos posteriores (Scherre, 1997; Naro e Scherre, 1993; 2007), a autora não utilizou algumas variáveis, ao que parece por não apresentarem clareza quanto aos resultados. Uma variável estrutural que se mostrou confusa, por exemplo, foi a variável pluralidade do contexto, em que a autora propôs investigar se um contexto de maior pluralidade nominal provocaria a existência de SNs totalmente marcados e, inversamente, se um contexto de menor pluralidade provocava a existência de SNs parcialmente marcados. Os resultados foram um tanto interessantes ao mostrarem que o contexto favorecedor da CN foi o contexto posterior ao SN, contrariando o princípio das marcas precedentes, também denominado como paralelismo formal (NARO e SCHERRE, 1993).

As variáveis sociais consideradas no estudo de Scherre (1988) para a comunidade analisada revelam que: (i) quanto maior a escolaridade do falante, maior também o emprego da CN; (ii) as mulheres são as que mais utilizam as marcas de plural, realizando mais o padrão de aplicação da CN; e (iii) a faixa etária mostrou um padrão de variação estável devido à ação de outros fatores, como exposição à mídia, mercado ocupacional e sensibilidade linguística.

Scherre (1988) ainda acrescenta que, de maneira geral, a concordância não é uma necessidade lógica das línguas, visto que a marca de plural formal ou semântica em quaisquer constituintes do sintagma já é suficiente para dar a ideia de pluralidade. No caso do PB, a tendência geral é marcar preferencialmente a primeira posição do SN e, a partir disso, as demais marcas não são necessárias para o entendimento da informação.

### **2.2.2 A concordância nominal no português do Estado da Bahia**

Para elucidar a diversidade de estudos sobre a CN em diversas variedades do PB e, posteriormente, estabelecer uma comparação entre eles, apresento nesta seção uma síntese de importantes trabalhos realizados com amostras de fala do português do Estado da Bahia: Lopes (2001), Andrade (2003, 2006) e Baxter (2009). Saliento que a maioria das variáveis elencadas nesses estudos são as mesmas utilizadas no estudo de Scherre (1988), por isso, dou maior destaque aos resultados das variáveis sociais para mostrar a existência de um quadro de variação estável ou mudança em curso, bem como menciono a possível relação dos resultados obtidos nessas

pesquisas com o contato linguístico ocorrido nas comunidades de fala estudadas, conforme indicam os referidos autores.

### 2.2.2.1 O português de Salvador

Lopes (2001) analisou a variação da CN em dados de fala popular e culta do português da cidade de Salvador. A amostra contém 74 inquéritos compostos por dados do Projeto NURC (Projeto Norma Culta Urbana) da década de 1970 e de 1990 e dados do PEPP (Programa de Estudos sobre o Português Popular de Salvador)<sup>4</sup>. O objetivo principal da comparação entre as gerações foi verificar se a variação em relação ao uso da CN é estável ou se se trata de um processo de mudança em curso. Além disso, Lopes (2001) analisa se o português de Salvador reflete características de processo de aquisição diferentes de outras regiões.

Os 74 inquéritos analisados por Lopes (2001) englobam 12 informantes do Projeto NURC/1970, 48 informantes do PEEP e 18 informantes do NURC/1990. A autora fez uma análise mórfica dos dados, observando cada elemento do SN e adotando a CN adotando a presença *versus* ausência da marca de plural como variável dependente. As variáveis linguísticas consideradas por Lopes (2001) foram: saliência fônica (processos de formação de plural, tonicidade), posição linear, categoria morfológica, posição relativa ao núcleo, grau dos substantivos e adjetivos, marcas precedentes, contexto fonológico subsequente e coexistência com o *tudo* (quantificador). As variáveis sociais foram: escolaridade, gênero, faixa etária, tempo, etnia.

Os resultados dos fatores estruturais encontrados por Lopes (2001) apontam similaridades quanto aos resultados com as demais variedades observadas do PB, como a do Rio de Janeiro estudada por Scherre (1988). No entanto, alguns cruzamentos entre as variáveis estruturais significativas e as variáveis sociais, no estudo de Lopes (2001), apresentaram resultados peculiares em relação à comunidade de fala de Salvador.

---

<sup>4</sup> O Projeto NURC foi iniciado em 1969, através do trabalho conjunto de vários pesquisadores de universidades brasileiras, com o propósito de descrever a fala de indivíduos de escolaridade superior em cinco capitais do país: Recife, Salvador, Rio de Janeiro, São Paulo e Porto Alegre. O PEPP é um projeto conjunto da UNEB (Universidade do Estado da Bahia) e da UFBA (Universidade Federal da Bahia), o qual tem por objetivo dar condições de análise diversificada acerca da fala popular de Salvador.



Para a relação da saliência fônica com a escolaridade, os resultados mostram o seguinte: 1) os falantes universitários mais velhos são os que mais fazem a CN independentemente do nível de saliência fônica dos vocábulos; 2), a oposição mais saliência, mais concordância *versus* menos saliência, menos concordância ocorre em todos os grupos de informantes, como mostram os exemplos (32) e (33), em que há dados de fala dos menos escolarizados e dos falantes universitários, respectivamente.

(32) a. muitos JOGOS eletrônico  
b. algumas COISA ruim

(33) a. dez REAIS era o quê?  
b. minhas IRMÃ pra praia

(LOPES, 2001, p.214)

Como se vê em (32a), a palavra *jogos* apresenta o morfema de plural *-s*, essa palavra possui uma oposição mais perceptível em relação à forma singular *-jogo/jogos*. A marcação de plural é favorecida pelo fato de a oposição de número implicar em mudança sonora, o que não ocorre na oposição coisa/coisas, apresentada em (32b). A marcação de plural também é favorecida em itens que possuem a forma singular terminada em *-l* e a forma plural em *-is*, como se vê em (33a), as formas *real/reais* apresentam diferenciação fônica na oposição singular/plural. Isso não ocorre em *irmã/irmãs*, exemplificada em (33b), cuja diferenciação fônica é menos perceptível. Assim, os falantes tendem a marcar mais o plural nas formas mais salientes.

A análise da interação da saliência fônica com a variável etnia revela que os descendentes de africanos marcam mais o plural quando a diferenciação morfofonológica entre as formas singular e plural é mais perceptível. A variável etnia teve como base a utilização da origem dos sobrenomes dos informantes, colocados em dois grupos: sobrenomes religiosos (informantes descendentes de africanos) e sobrenomes não-religiosos (descendentes de outras etnias).

Embora Lopes (2001) tenha se dedicado em controlar as características étnicas dos informantes, adotar a origem de seus sobrenomes para distingui-los como afrodescendentes ou não deixa lacunas, pois a diversidade étnica do Brasil sugere uma certa imprecisão para definir esse aspecto. A autodeclaração étnico-racial parece uma alternativa mais viável para controlar a variável etnia.

Lopes (2001) diz que pode haver uma relação mais direta da fala dos descendentes de africanos com um passado de aquisição do português no que diz respeito à percepção dos dados mais salientes. Segundo a autora, no processo de aquisição, os falantes adquirem mais rapidamente as formas salientes, enquanto as formas menos salientes ficam desobrigadas de participarem do processo porque a ausência de marcação de plural nesses é menos perceptível. Entretanto, a saliência fônica parece não seguir essa pressuposição no estudo de Baxter (2009) sobre o português afro-brasileiro de Helvécia e sobre o português dos Tongas (África), visto que essa variável não foi significativa na análise dos dados. Conforme veremos no Capítulo 5, neste trabalho a saliência fônica também apresentou um resultado distinto do que mencionou Lopes (2001): os SNs com itens mais salientes receberam menos concordância.

O efeito da relação entre posição linear e classe gramatical nos níveis diferentes de escolaridade foi semelhante: os itens localizados à esquerda do núcleo receberam mais marcação de plural do que os itens pós-nucleares, como o exemplo (34) em que o pronome demonstrativo *essas* se encontra na primeira posição do SN e recebe a marca de plural.

(34) “ESSAS coisa; eu faço pra mim” M3C17 p. 244

(LOPES, 2001, p. 245)

Sobre a interação entre a posição linear do constituinte e a etnia, os resultados mostraram que, nos dois grupos de sobrenome, a oposição entre posição à esquerda do núcleo mais favorecedora. A posição à direita foi menos favorecedora da aplicação da CN se manteve nos dois grupos étnicos, conforme exemplo (35), o constituinte em primeira posição recebe a marca de plural, enquanto os demais constituintes não recebem.

(35) “aquelas camisa BRANCA, com letreiro aqui” M1F05

(LOPES, 2001, p. 245)

As variáveis marcas precedentes, contexto fonético/fonológico subsequente e grau dos substantivos e adjetivos apresentaram neutralidade no tocante aos fatores sociais, o que evidencia que os seus resultados possuem relações com aspectos estruturais.

A única variável linguística analisada por Lopes (2001) que não consta no estudo de Scherre (1988) é a coexistência com o *tudo* (quantificador), usado também no valor de *todos*, conforme o exemplo (36).

(36) “as menina TUDO chutando, pegando” H3C15

(LOPES, 2001, p. 328)

O quantificador *tudo* possui um valor semântico de plural, por isso pode ser utilizado com o valor de *todos/todas* pelos falantes. Os resultados dessa variável no trabalho de Lopes (2001) mostram que a concordância entre os outros elementos do sintagma é desfavorecida quando o *tudo* com valor de *todos/todas* está presente, como podemos ver em (36), enquanto os sintagmas que sem o *tudo* se mostraram neutros quanto a aplicação da CN.

Em relação ao nível de escolaridade, não foi encontrado nenhum dado de *tudo* com valor de *todos/todas* nos dados dos falantes universitários. Ao analisar somente os falantes do português popular, dos níveis fundamental e médio, Lopes diz que o efeito dessa forma sobre a CN não está diretamente ligado à pouca escolaridade, uma vez que a sua presença afeta mais a concordância nos elementos do sintagma quando os falantes possuem mais escolaridade do que com menos (LOPES, 2001, p. 338).

No que tange à variável etnia, o quantificador *tudo* apresenta relação com a não aplicação da regra de CN na fala dos informantes descendentes de africanos. Lopes (2001) diz que esse resultado parece estar mais ligado a uma história de aquisição linguística com diferentes dados em relação a outros grupos, e que, em alguma fase da sua história de língua, deve ter havido o uso geral do quantificador sem concordância na forma do *tudo*. A autora acrescenta ainda que uso desse quantificador, no português de Salvador, é comparável ao que se encontra em comunidades como afro-brasileiras, nas quais houve uma história de contato de línguas e aprendizagem de português como primeira (L1) ou como L2 com muita diversidade e variação nos dados e sem muita força de controle e padronização (p. 342-343).

Analisando apenas os efeitos das variáveis sociais no uso da CN no português de Salvador, Lopes (2001) destaca a escolaridade, pois foi variável social que mais interagiu com as variáveis estruturais da CN e com as outras variáveis sociais. Os

resultados mostram que quanto mais tempo o indivíduo tem de contato com o ambiente escolar, maior a probabilidade de se fazer a concordância.

A variável gênero aponta que as mulheres possuem uma leve tendência a realizar mais a CN do que os homens. Cruzando os dados dessa variável com a escolarização, Lopes (2001) constatou que, entre os informantes de escolaridade média, as mulheres apresentam um maior percentual de aplicação da CN do que os homens. Lopes infere que as mulheres desse grupo, em Salvador, sofrem mais pressão social justamente por apenas possuírem o segundo grau escolar e, conseqüentemente, serem colocadas aquém dos homens no mercado de trabalho.

A faixa etária mais velha, informantes com mais de 65 anos, é a que apresenta maior percentual de marcação de plural. Segundo a autora, esse resultado parece indicar uma mudança da perda das marcas explícitas de plural, resultado contrário à hipótese deste trabalho para o português do interior da Bahia, cuja hipótese é de que os falantes mais jovens apresentem maior aplicação da regra (cf. Capítulo 5, subseção 5.3.2). Analisando conjuntamente as variáveis faixa etária e gênero, Lopes afirma que as mulheres jovens marcam mais o plural do que os homens da mesma idade.

Quanto à faixa etária ligada à escolaridade, Lopes (2001) verificou que, nos grupos de escolaridade fundamental e superior, há uma tendência ao aumento da aplicação da CN dos mais jovens para os mais velhos. Um ponto interessante que destaque no trabalho de Lopes (2001) sobre essas variáveis é que tanto a ausência como a presença de CN coexistem de acordo com o avanço da faixa etária: os falantes jovens do ensino fundamental apresentaram um quadro maior de ausência de concordância do que os mais velhos do mesmo nível de escolarização; no grupo dos falantes de nível médio, acontece o oposto, os mais novos fazem mais concordância do que os mais velhos, fato que pode ser interpretado como um processo de aquisição da regra.

No grupo dos falantes de nível superior da cidade de Salvador, a aplicação da regra de CN vai aumentando conforme o avanço da idade. Lopes (2001) sugere que isso tem relação com o fato de que esse grupo de pessoas estudou numa época em que somente quem possuía boas condições financeiras ingressava na educação superior. Já os universitários mais jovens realizam menos a CN, o que pode ser configurado como uma tendência à perda das marcas de plural.

Para verificar a possibilidade de mudança, Lopes (2001) comparou esses resultados com dados do NURC/1970 e NURC/1990. Nessas amostras, ela observou

que o comportamento dos falantes mais jovens e dos mais velhos indica uma mudança geracional e não uma mudança em progresso, pois, em ambas as sincronias, os mais jovens aplicam menos a CN do que os mais velhos. Quanto à fala dos universitários, na comparação das sincronias, Lopes constatou uma leve redução na aplicação da CN conforme o passar do tempo.

A variável etnia mostrou que os informantes com sobrenomes religiosos fizeram menos o uso da norma padrão da CN do que aqueles com sobrenomes não-religiosos. Ao analisar separadamente os dados dos informantes dos níveis fundamental e médio, Lopes (2001) verificou que os informantes considerados descendentes de africanos, os de sobrenome religiosos de origem negra, aplicaram menos a regra de CN, ou seja, há uma relação entre esse fenômeno linguístico e a etnia.

Explorando separadamente os grupos de sobrenome, em todas as faixas etárias, e analisando a relação entre a faixa etária e a concordância, chega-se a um resultado interessante [...] esses resultados indicam que os dois grupos revelam situações diferentes, em relação ao fenômeno da concordância. Os informantes de sobrenome religioso, aqui considerados como de descendência negra, estão adquirindo a regra de concordância, ao invés da perda: os mais jovens estão fazendo mais concordância que os mais velhos, o que parece indicar um processo de mudança em andamento. (LOPES, p. 187).

Ao comparar os resultados de sua pesquisa com outras realizadas anteriormente, Lopes concluiu que a variação da CN no PB, além das restrições linguísticas, tem como um fator bastante relevante a escolaridade, ou seja, confirma que “quanto mais escolaridade, mais concordância, demonstrando que o trabalho institucional e sistemático da escola, que busca a manutenção do padrão, ainda funciona” (LOPES, 2001, p. 381). Ela ainda faz uma importante afirmação sobre o português da cidade de Salvador:

Como em Salvador a formação da população constituiu-se de um confronto de grupos com história de diferentes tipos de aquisição do português, com mais e menos variação, a variação da concordância nesta cidade apresentou reflexos desse tipo de aquisição. Apesar de serem detectadas diferenças com relação às restrições na variação da concordância no sintagma nominal, entre os diversos grupos não foram constatadas evidências de que as gramáticas são diferentes. Ao separar pessoas de diferentes etnias, nota-se, contudo, que enquanto os descendentes dos escravos cada vez demonstram fazer mais concordância, em um processo de aquisição da regra, os não descendentes dos escravos, em processo inverso, tendem a fazer menos concordância. (LOPES, 2001, p. 381).

### 2.2.2.2 O português rural afro-brasileiro baiano

Andrade (2006), com base em *corpus* do português afro-brasileiro, do Projeto Vertentes, analisou a CN em três comunidades rurais baianas formadas por falantes descendentes de africanos: Helvécia, Sapé e Rio de Contas. As três comunidades estão localizadas em diferentes microrregiões do Estado da Bahia, mas são semelhantes por terem permanecido isoladas durante muito tempo, fato que pode ter levado seu dialeto a conservar algumas peculiaridades típicas do português L2 com muitas marcas de línguas africanas (cf. ANDRADE, 2003; LUCCHESI, 2009).

O estudo de Andrade (2006) considerou a perspectiva de análise sintagmática e teve como variável dependente a aplicação *versus* a não-aplicação da regra de CN, em que o valor positivo é atribuído a SNs cujos constituintes são todos marcados (*os meninos*) e, àqueles SNs em que nem todos os elementos recebem marcação de plural é atribuído o valor negativo (*os menino*). A autora teve como objetivo principal verificar a influência do contato linguístico no português falado nessas comunidades descrevendo os resultados das variáveis configuração sintagmática do SN e saliência fônica dos elementos do SN em conjunto com as variáveis sociais faixa etária, escolaridade e estada fora da comunidade e idade.

Os resultados gerais apresentam um uso muito baixo da regra de concordância nos dialetos de Helvécia, Rio de Contas e Sapé (8%), conforme já era esperado pela autora. Ela afirma que esse resultado está relacionado à proximidade dos dialetos com a gramática do português L2 falado anteriormente nas comunidades, o qual possuía marcas de línguas africanas cuja tipologia diferente do português apresentava marcação de concordância distinta. Andrade (2006) ainda diz que, embora o percentual de aplicação da regra de CN no SN seja baixo, ele é presente no dialeto das comunidades e evidencia a ação de fatores linguísticos e sociais.

A variável configuração sintagmática do SN mostra que a estrutura do SN formada por quantificador + nome é que mais favorece a marcação de CN, como no exemplo (37), seguida das estruturas que possuem um numeral, exemplificado em (38). De acordo com Andrade (2006), talvez a concordância seja favorecida pela presença do numeral porque possui transparência semântica, enquanto outros SNs cujos elementos são pluralizados com a inserção do morfema funcional -s possuem uma representação mais abstrata de plural.

(37) Já tem três. **Todos três homens.**

(38) Ah, tem **uns... uns vinte e dois anos** mais ou menos, que...

(ANDRADE, 2006, p. 2077)

Segundo a autora, o dialeto de Helvécia, em particular, possui algumas características quanto à estrutura morfossintática: a maioria dos SNs apresenta a estrutura de determinante + núcleo (nome), sendo os determinantes em sua maioria artigos, pronomes possessivos ou demonstrativos, e os núcleos formados por substantivos, exemplo (39). Neste trabalho, a configuração sintagmática dos SNs revelou estruturas morfossintáticas semelhantes à Helvécia (cf. Capítulo 5, subseção 5.2.1).

(39) As pessoa de lá é muito bom [Helvécia-M3]

(BAXTER, 2009, p. 270)

Andrade (2003) relacionou essa simplificação da estrutura morfossintática do SN, bem como a maior frequência de dados sem marcas de plural com processos históricos de formação da comunidade que remontam para um passado em que as marcas de concordância tendiam a ser menores. Os resultados obtidos na análise da variável faixa etária sugerem isso, ao mostrarem que os informantes mais velhos usam menos as marcas de plural do que os mais jovens, resultado contrário ao de Lopes (2001) sobre o português de Salvador.

Morfemas que têm uma carga de plural mais transparente contribuem para a aplicação da regra de concordância, especialmente em situações de contato linguístico, em que as formas mais salientes são primeiramente adquiridas (GUY, 1981; LOPES, 2001; ANDRADE, 2003; ANDRADE, 2006). Mesmo não seguindo a hierarquia de saliência (SCHERRE, 1988), os resultados do estudo de Andrade indicaram que a CN foi mais empregada quando o SN apresentou itens mais salientes, como em (40). Na subseção 5.2.5, descrevo os resultados da saliência fônica no português do interior da Bahia, os quais apontam comportamento distinto do estudo de Andrade (2006).

(40) **Os eleitores** dele que se exhibe.

(ANDRADE, 2006, p. 2079)

A análise da variável faixa etária mostra que a regra de CN no SN está em processo de mudança aquisicional nas comunidades afro-brasileiras rurais baianas, pois os falantes mais jovens estão realizando mais concordância do que os mais velhos. Os resultados da variável estada fora da comunidade revelaram que o contato com pessoas de fora da comunidade, bem como o convívio fora da comunidade contribuem significativamente para o processo de aproximação do dialeto aos padrões de maior prestígio social.

Quanto à escolarização, os informantes das comunidades foram divididos entre aqueles que possuíam pouca escolaridade, semianalfabetos, e aqueles que não tinham nenhuma escolaridade, analfabetos. Os resultados foram um tanto inesperados, os informantes sem escolaridade alguma realizando mais a regra do que os que frequentaram a escola por algum período de tempo. Apesar de ir de encontro com o padrão, a análise dessa variável reflete, segundo Andrade (2006), a realidade das comunidades estudadas porque a escola não exerce influência no desenvolvimento dialetal dos adultos, visto que muitos dos informantes da amostra que foram classificados como semianalfabetos não tiveram um ano sequer de estudo.

Podemos levar em conta, também, que até mesmo as instituições públicas de ensino da zona urbana funcionam precariamente e com um quadro de profissionais, muitas vezes, mal formados ou até mesmo leigos que não dominam de forma efetiva a variedade culta da língua portuguesa, não podendo, dessa forma, ampliar o universo dialetal de seus alunos, e, na zona rural esse quadro é ainda mais grave. Mesmo que se encontrem, em zonas afastadas, profissionais capacitados, não podemos obscurecer o fato de que muitos alunos não conseguem aprender, simplesmente porque não entendem a variedade lingüística de seu professor (ANDRADE, 2006, p. 2081).

Baxter (2009) apresenta uma análise mórfica da CN comparando duas comunidades afrodescendentes: uma brasileira, a comunidade de Helvécia, já mencionada em Andrade (2003; 2006), e uma africana, a comunidade dos Tongas, da roça Monte Café, em São Tomé e Príncipe.

A escolha dessas duas comunidades, segundo o autor, foi motivada pela semelhança de fatos sócio-históricos e lingüísticos na formação: são variedades de português que surgiram em micro sociedades de trabalho controladas e fechadas, situação de escravatura; são oriundas de contextos demográficos com altas



proporções de africanos e afrodescendentes contratados como administradores; seus antepassados africanos aprenderam o português pelo contato a partir de modelos falados pelos colegas trabalhadores e pelos administradores, originando um contato linguístico do português com línguas africanas, fato que pode estar relacionado com fenômenos linguísticos variáveis, como a CN nos exemplos (41) e (42).

- (41) a. As peessoa de lá é muito bom [Helvécia-M3]  
 b. Bota duas culeres n'agua morna [Helvécia-M2]
- (42) a. Os outros disse [Monte Café-H3]  
 b. Issos coesa tudo [Monte Café-H3]

(BAXTER, 2009, p. 270).

Baxter (2009) considerou como variável dependente para a análise mórfica a ausência *versus* a presença de marca de plural em cada item do SN separadamente para analisar a CN na fala de 36 informantes (18 de cada comunidade), divididos em quatro faixas etárias: faixa 1: de 21 a 40 anos; faixa 2: de 41 a 60 anos; faixa 3: de 61 a 80 anos; e faixa 4: de mais de 85 anos. O autor apresentou os resultados das variáveis linguísticas posição linear do item em relação ao núcleo e função, e marcas precedentes. Com relação às variáveis sociais, Baxter (2009) observou a faixa etária, o gênero e a estada fora da comunidade.

De modo geral, a comparação dos resultados dos dados de Helvécia e dos Tongas apresentou semelhança nos papéis da estrutura pré-nuclear, nuclear e pós-nuclear na marcação de plural, bem como na análise independente de cada faixa etária. Baxter (2009) afirmou que, em todas as faixas etárias, há um declínio no grau de marcação do morfema de plural na posição da esquerda para a direita do SN, isto é, os informantes das duas comunidades marcam mais o plural na posição pré-nuclear do que no núcleo ou nas posições pós-nucleares. Nas faixas etárias mais avançadas, isso ocorreu com maior frequência, conforme se pode observar no Gráfico 1<sup>5</sup>.

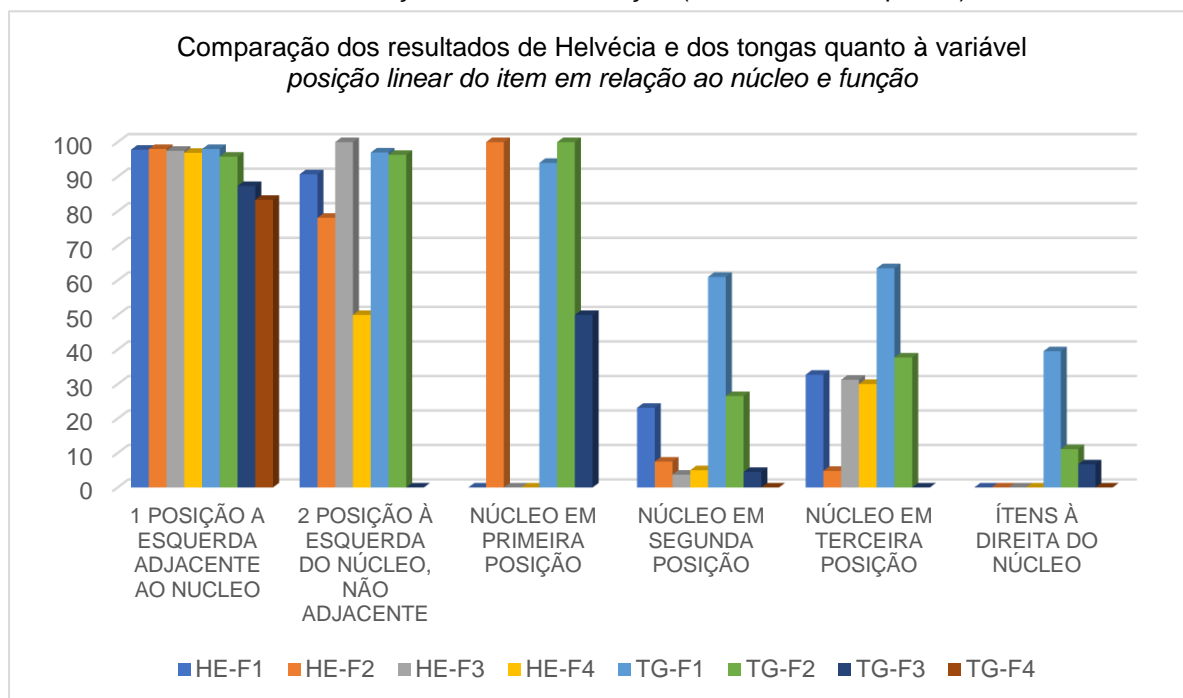
<sup>5</sup> As abreviaturas contidas na legenda do Gráfico 1 correspondem a:

HE – Helvécia

TG – tongas

F1 – faixa etária 1 (20 a 40 anos); F2 – faixa etária 2 (41 a 60 anos); F3 – faixa etária 3 (61 a 80 anos); F4 – faixa etária 4 (85 a 103 anos).

Gráfico 1. Comparação dos resultados de Helvécia e dos Tongas quanto à variável posição linear do item em relação ao núcleo e função (BAXTER, 2009, p. 278).



Fonte: elaborado pela autora com base nos dados de Baxter, 2009, p. 278.

Quanto à variável marcas precedentes, as comunidades mostraram um comportamento diferente. Em Helvécia, o princípio do paralelismo formal não é relevante porque: (i) a presença de marca formal na primeira posição não favorece a marcação do segundo item; (ii) a marcação do item em terceira posição é desfavorecida, tanto pela presença de uma mistura de marcas com marca formal precedente como pela presença da mistura de marcas com zero formal precedente.

Já no dialeto dos Tongas, esse princípio se revela contraditório: a marca flexional precedente na primeira posição leva à marca na segunda posição, e mistura de marcas precedentes com marca formal levam à marca na terceira posição, no entanto, zero em primeira posição favorece a marcação da segunda posição, enquanto a mistura de marcas precedentes com zero favorece zero na terceira posição. Os resultados das duas comunidades podem ser vistos no Quadro 2.

Quadro 2. Resultados da variável marcas precedentes nos dialetos de Helvécia e dos Tongas.

Marcas precedentes	Helvécia Análise Global (quatro faixas etárias)			Tongas (Monte Café) Análise Global (quatro faixas etárias)		
	Nº de dados	%	PR	Nº de dados	%	PR
Marca formal na 1ª posição; item analisado na 2ª posição ( <i>dos meus pai</i> )	77/1031	7.5	.499	481/1001	48.1	.583
Numeral na 1ª posição; item analisado na 2ª posição ( <i>dez filhos</i> )	58/425	13.6	.641	194/936	20.7	.413
Ausência de marca formal na 1ª posição; item analisado na 2ª posição ( <i>do meus pai</i> )	10/18	55.6	.884	16/45	35.6	.651
SN de três itens: mistura de marcas precedentes com marca formal; item analisado na 3ª posição ( <i>seus quatro meses</i> )	8/65	12.3	.412	68/130	52.3	.557
SN de três itens: mistura de marcas precedentes com zero; item analisado na 3ª posição ( <i>toda essas coesa</i> )	3/113	2.7	.099	3/44	6.8	.190

Fonte: Adaptado de Baxter (2009, p. 281)

O princípio do paralelismo formal parece marginal à atribuição de plural nesses dois dialetos que manifestam estruturas de SN e padrões de marcação de plural muito simples, segundo Baxter (2009). Ele ainda faz uma consideração importante: diferentes dialetos podem desenvolver diferentes padrões de marcação no que diz respeito à influência das marcas precedentes, embora com semelhanças parciais (BAXTER, 2009, p. 282).

A faixa etária é a variável social que indica a inclinação das comunidades para a aquisição das marcas de plural, pois em ambas as comunidades, a faixa etária mais avançada apresentou um número bem reduzido de marcação de plural no SN comparado à faixa etária mais jovens.

A variável gênero teve resultado similar nas duas comunidades: os homens marcam mais o plural nos SNs do que as mulheres. Esse resultado está diretamente ligado ao resultado da variável estada fora da comunidade, pois são os homens que passam mais tempo fora da comunidade a trabalho os homens têm um maior grau de integração social, o que implica um maior contato com padrões linguísticos adventícios, com a norma linguística de prestígio.

Os resultados das variáveis sociais são indicativos de um perfil aquisicional quanto à estrutura da variável de plural no dialeto de Helvécia e dos Tongas. Este estudo comparativo entre uma comunidade afro-brasileira isolada e uma comunidade africana falante de português feito por Baxter (2009), bem como o estudo de Andrade (2006) com as comunidades afro-brasileiras rurais contribuem para a observação das semelhanças e das diferenças sobre o comportamento da CN na configuração estrutural do PB no eixo rural-urbano, pois as variedades de português faladas por afrodescendentes em contextos rurais apresentam especificidades, em virtude de suas características sócio-históricas, distintas das variedades urbanas.

Essa evidência reafirma uma das propostas deste trabalho que é verificar a existência de um *continuum* linguístico da CN que perpassa as regiões de Poções e de Santo Antônio de Jesus, as quais compõem o *corpus* do português do interior da Bahia.

### 2.3 SÍNTESE

Considerando os objetivos propostos na seção 2.1, elenco a seguir as informações relevantes para:

1) As resenhas dos estudos de Andrade (2003, 2006), Baxter (2009) e de Lopes (2001) possibilitaram a fixação das variáveis consideradas neste trabalho: configuração sintagmática do SN, realização do núcleo, número de constituintes do SN, número de constituintes flexionáveis do SN, saliência fônica, classe gramatical do constituinte marcado, posição linear do constituinte marcado, sexo, faixa etária, escolaridade, estada fora da comunidade e localidade. A análise dessas variáveis viabiliza a verificação da existência de um *continuum* entre o português rural afro-brasileiro e o português de Salvador, pois a aplicação da regra de concordância naquele é menor (8%) do que nesse (34%);

2) O comportamento morfossintático da CN no português do interior da Bahia tende a ser semelhante a outras variedades do PB no que tange aos fatores linguísticos, pois os estudos de Scherre (1988) para o Rio de Janeiro, Lopes (2001) para Salvador, Andrade (2003, 2006) e Baxter (2009) para comunidades afrodescendentes apresentam semelhanças pontuais. Nesse sentido, esses estudos são relevantes, pois norteiam este trabalho quanto ao fornecimento de uma matriz teórico-metodológica;

3) Nas variedades do PB apresentadas nas resenhas dos trabalhos, o quadro de variação na aplicação da regra de CN encontra-se em variação estável no Rio de Janeiro (Scherre, 1988) e em Salvador (Lopes, 2001). Porém, nas comunidades rurais afrodescendentes (Andrade, 2003, 2006; Baxter, 2009), a aplicação da CN encontra-se em processo de mudança em curso.

Diante do exposto neste Capítulo, percebo um dos grandes desafios para quem estuda linguística: sistematizar aquilo que é dado como inacabado e que está em evolução, em mudança. Por isso, é importante recorrer aos estudos anteriores para resgatar ideias, valores e proposições que podem contribuir para o aprofundamento sobre o objeto de estudo do linguista, especialmente, do sociolinguista: a língua em uso real.

### 3 O CONTATO LINGUÍSTICO NA FORMAÇÃO DO PORTUGUÊS BRASILEIRO

Os aspectos teórico-metodológicos nos quais se debruça este trabalho de pesquisa consideram que a variação e a mudança linguística são impulsionadas não apenas por fatores internos às línguas, mas também por fatores socioculturais resultantes de processos históricos (WEINREICH, LABOV E HERZOG, 2006[1968]; MATTOS E SILVA, 2004a; NEGRÃO E VIOTTI, 2012).

Por esse motivo, é indispensável fazer uma sistematização sobre alguns fatos históricos acerca da formação do povo brasileiro e do PB, sobretudo, sobre as características do português popular do interior da Bahia, o qual forma o *corpus* deste trabalho.

Nesse capítulo, o objetivo centra-se, portanto, em: (i) refletir sobre aspectos sócio-históricos e sociolinguísticos, apresentando as principais teorias do contato linguístico e seu papel na formação de novas variedades de língua; (ii) discorrer sobre a participação das matrizes étnicas dos portugueses, indígenas e africanos e de suas línguas no período colonial do Brasil; (iii) expor brevemente o processo de urbanização do Brasil, considerando a proposição de Bortoni-Ricardo (2005) sobre um *continuum* do PB e a proposta de Lucchesi (2002, 2015) que reconhece o modelo do *continuum* e o modelo da polarização do PB como complementares.

#### 3.1 TEORIAS SOBRE A AQUISIÇÃO DE LÍNGUA EM SITUAÇÃO DE CONTATO

Estudar a mudança linguística, bem como a formação de novas línguas a partir da perspectiva do contato linguístico, tem sido a motivação de diversas pesquisas, as quais seguem ou refutam diversas hipóteses linguísticas. A maioria das pesquisas investigam aspectos sobre *pidgins* e crioulos, variedades de línguas novas formadas a partir de contextos sócio-históricos específicos que diferem das línguas que passam por mudança linguística considerada regular.

As hipóteses focalizadas nesta seção apresentam como principal característica distintiva o fato de serem ou não excepcionalistas. As hipóteses excepcionalistas defendem que as línguas crioulas possuem basicamente as mesmas características

e fazem distinção entre línguas não crioulas e crioulas, atribuindo-lhes um *status* de “simplicidade”. São hipóteses excepcionalistas: *Bioprograma da Linguagem* (BICKERTON, 1988); *Relexificação* (LEFEBVRE, 1997, 1998, 2001; LEFEBVRE E LUMSDEN, 1994); *Transmissão Linguística Irregular* (BAXTER, 1992, 1995; BAXTER e LUCCHESI, 1997; LUCCHESI e BAXTER, 2009). Com base no que diz DeGraff (2003), identifico nessas teorias uma visão excepcionalista sobre as línguas crioulas porque generalizam o processo de aquisição como imperfeito, no qual as peculiaridades da nova língua que se forma seriam perdas gramaticais.

De maneira distinta, as hipóteses não-excepcionalistas são aquelas que propõem o estudo das línguas crioulas a partir dos aspectos sócio-históricos que envolvem situações de contato linguístico e formação de novas línguas, tendo em vista que a principal diferença entre as línguas crioulas e as não crioulas é a formação sócio-histórica (ABOH, 2009). São hipóteses não-excepcionalistas: *Competição e seleção por ecolinguística* (MUFWENE, 2001, 2002, 2008); *Teoria Nula* (ABOH, 2015; ABOH E DEGRAFF, 2017); e outras proposições sobre formação crioula e aquisição de L2 (BAPTISTE, 2016).

Outro aspecto relevante dessas é a perspectiva não genérica sobre os efeitos do contato, considerando o estudo mais aprofundado das propriedades morfossintáticas das línguas. Nesse sentido, parte-se do princípio de que o processo de aquisição da L2 é ancorado no conhecimento das propriedades gramaticais das línguas que estiveram em contato e não em uma possível dificuldade dos aprendentes em dominar a LA.

Nas subseções de 3.1.1 a 3.1.6, apresento os principais aspectos das hipóteses excepcionalistas e não-excepcionalistas que consideram o contato na aquisição e na mudança linguística. Antes, porém, exponho os conceitos e a caracterização sócio-histórica de línguas *pidgins* e crioulas.

Vários autores, dentre eles Baxter (1996), Holm (1998), Hagemeyer (2009), dizem que a principal característica sócio-histórica da formação de *pidgins* e crioulos é a situação em que grupos de adultos falantes de diferentes línguas são postos em contato e precisam interagir, ou seja, a necessidade de comunicação emergencial leva esses falantes a desenvolverem uma espécie de código, denominado por alguns estudiosos como jargão ou *pré-pidgin*, que possui regras gramaticais das línguas nativas e vocabulário misto. Assim, de acordo com Baxter (1996), restrições sociais

impedem o processo de aprendizagem de uma L2 e nenhum grupo adquire a língua do outro grupo.

Para compreender o que são e como se formam as línguas *pidgins* e crioulas, é fundamental compreender os fatores sócio-históricos que caracterizam a formação e o desenvolvimento dessas línguas. O primeiro fator a ser considerado é o contexto de multilinguismo, em que, geralmente, há uma língua de superstrato<sup>6</sup>, falada por uma pequena elite, e várias línguas de substrato, faladas por uma maioria subserviente. Desse contato pode resultar uma língua franca<sup>7</sup>.

Devido à necessidade de comunicação, surge um jargão ou *pré-pidgin*, código linguístico para comunicação inicial e emergencial caracterizada por apresentar uma grande redução gramatical. Se utilizado por determinado tempo, mediante a uma situação de contato linguístico mais duradoura, ocorre o que se chama de *pidginização*. Conforme Baxter (1996), a língua que começa a se formar passa a ser um *pidgin*, na qual há uma redução do léxico e de aspectos funcionais gramaticais. Além disso, o que caracteriza a pidginização é o fato de a nova variedade de língua ser utilizada pelos falantes do substrato para a interação entre si e não somente como forma de comunicação com o dominador (BAXTER, 1996; COUTO, 1996; LEFEBVRE, 1998).

Segundo esses autores, quando um pidgin se nativiza<sup>8</sup>, há o surgimento de uma língua crioula, isto é, quando uma nova geração de falantes adquire como língua materna a variedade de língua formada nessa comunidade em função do contato. A língua crioula passa a ser a língua materna da nova comunidade de falantes, mesmo que esses tenham acesso às línguas nativas dos pais, e é caracterizada por ser mais difundida e estendida no léxico e na gramática.

---

<sup>6</sup> Língua de substrato – nome que se dá à língua de um povo que é abandonada em proveito de outra que a ela se impõe, geralmente como consequência de uma conquista política; língua de superstrato – nome que se dá à língua de um povo conquistador, que a abandona para adotar a língua do povo vencido (CÂMARA JÚNIOR, 1981, p. 227-228).

<sup>7</sup> É a língua tomada como língua comum de grupos sociais que falam, cada um, uma língua diferente dos outros. Por exemplo, em países africanos de colonização Portuguesa, os diversos povos organizados num mesmo País, falam cada um uma língua africana. Por outro lado, utilizam o português como língua para estabelecerem relações entre si. No Brasil, no período colonial, houve uma língua franca, a língua geral, que era falada pelos portugueses, tanto nas relações com índios como nas relações domésticas, e pelas diversas tribos indígenas, embora cada uma falasse a sua língua (*Enciclopédia das Línguas no Brasil, ELB*).

<sup>8</sup> 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.



De acordo com Couto (1996), as línguas crioulas são L1 para a maioria dos seus falantes, diferentemente dos *pidgins* que não têm falantes nativos. Conforme o mesmo autor, apesar de, ao longo da história das línguas naturais, ocorrerem muitos casos de pidginização devido aos contatos proporcionados pelos deslocamentos com interesse comercial e escravocrata, o primeiro *pidgin* relevante foi uma variedade do português utilizada para comunicação na costa africana e na rota marítima dos portugueses, utilizado para comunicação já desde os primeiros anos das Grandes Navegações.

Para a maioria dos estudos crioulísticos, as línguas crioulas de base europeia surgiram durante os séculos XV a XIX, principalmente no contexto da expansão ultramarina, quando ocorreu o maior fluxo de expansão europeia com a colonização de vários povos e a ampliação comercial. No entanto, hipóteses mais recentes discordam do status de língua crioula atribuído somente às línguas que se originaram do contato de línguas europeias com línguas africanas (MUFWENE, 2008; ABOH, 2015).

Os estudos sobre contato linguístico e formação de *pidgins* e crioulos têm buscado explicar a origem dessas línguas através de diversas hipóteses, algumas mais distintas do que outras, mas todas com o objetivo de discutir a respeito dessas línguas. Tendo em vista que o PB é uma língua formada sob condições sócio-históricas que corroboram um contato linguístico intenso e extenso, considero as proposições das hipóteses não-excepcionalistas coerentes com os objetivos do presente trabalho, sobretudo, o que propõem as hipóteses da *Competição e seleção por ecolinguística* (MUFWENE, 2001, 2002, 2008) e da *Teoria Nula* (ABOH, 2015; ABOH E DEGRAFF, 2017), conforme descrevo posteriormente.

### **3.1.1 Bioprograma da Linguagem**

A hipótese do Bioprograma da Linguagem de Derek Bickerton (1981; 1984; 1988) concentra-se na ideia que as línguas crioulas surgem de maneira abrupta ou catastrófica em oposição às línguas naturais que surgem e se formam de maneira gradual. Segundo Bickerton (1988), novas línguas surgem quando qualquer circunstância histórica faz com que um grupo de falantes originalmente multilinguístico abandone parcialmente ou completamente sua língua nativa, mas quando esse grupo

cria um código comunicativo emergente que mescla duas ou mais línguas, simplificando-as gramaticalmente, há a formação de *pidgins* e crioulos.

Além disso, Bickerton concentra seu estudo na perspectiva de que as crianças que nascem em situação de contato seriam os agentes principais na formação das línguas crioulas, afirmando que o *input* recebido para aquisição da língua materna seria um modelo de língua bastante deficiente de L2, o jargão ou pré-pidgin. Essas crianças acionariam os dispositivos da GU e preencheriam as lacunas gramaticais da L2 deficiente dos pais e, então, ocorreria uma reestruturação gramatical da língua crioula. Baseando-se nesse aspecto, Bickerton diz que as línguas crioulas seriam uma forma de janela para a GU, isto é, estudar essas línguas possibilitaria verificar como funciona a faculdade inata da linguagem.

Considerando a colonização europeia como favorecedora do surgimento de línguas crioulas, Bickerton definiu dois contextos sociolinguísticos que, segundo ele, caracterizam o processo de formação de *pidgins* e crioulos:

I) Um grupo de falantes de fora entra numa área multilinguística e se associa com o povo local, de forma permanente ou semipermanente, com o objetivo de estabelecer relações comerciais para obter poder, o grupo dominador se estabelece em locais onde já são faladas diferentes línguas e os dominados ainda são obrigados a aprender a língua do superstrato. Esse contexto foi denominado como *crioulos de forte* ou *situação de fort*,

II) Um grupo de falantes de fora cria uma nova sociedade transportando pessoas de diversas localidades para uma ilha ou uma área costeira inabitada ou sub-inabitada. Não havendo uma língua em comum entre eles, os trabalhadores escravizados são forçados a conviver e, por conseguinte, a comunicar-se. Esse contexto foi denominado de *crioulos de plantação* ou *situação de plantation*.

Apesar de mencionar esses dois contextos sócio-históricos e sociolinguísticos, Bickerton afirma que não há relação direta entre os eventos históricos e os processos linguísticos envolvidos na formação dos crioulos e que a definição das duas situações mencionadas anteriormente não exclui a possibilidade de outras circunstâncias terem ocorrido ou que venham a ocorrer futuramente.

Embora a *situação de fort* e a situação de plantação sejam distintas socialmente, havia pouca diferença nos processos linguísticos produzidos e reproduzidos nelas, conforme Bickerton (1988). O autor defende que o relevante nessas situações de contato é a tentativa de aquisição de uma língua europeia por

falantes não europeus. Como essa aquisição não ocorre de forma completa dada a aprendizagem defectiva, principalmente nos estágios iniciais, há a formação de uma nova variedade de língua. A característica principal da nova língua, conforme a hipótese do Bioprograma, é uma redução radical nas propriedades estruturais da LA original, tendo em vista a ideia de que as línguas crioulas possuem uma gramática mais simples do que as línguas que as originaram.

A hipótese do Bioprograma é excepcionalista porque, além de considerar as crianças como os principais agentes do processo de formação dos crioulos, pressupõe que os crioulos possuem a mesma gênese, sendo o contato da LA com as línguas não europeias o motivo principal da perda de um grande número de propriedades gramaticais, gerando a simplificação. Esse modelo de aprendizagem é baseado no léxico adquirido em parte pelos escravizados, os quais não dão conta de adquirir a variada gama de significados e funções dos itens lexicais da LA.

Conforme Bickerton, os escravizados adquirem apenas um subconjunto limitado do léxico e um subconjunto das propriedades lexicais e morfológicas que serviriam como modelo único para a aquisição das futuras gerações de falantes. Disso decorre, outra proposição do Bioprograma: a de que todos os crioulos possuiriam a mesma gênese devido à grande semelhança nos aspectos gramaticais, como a ausência de elementos gramaticais mais abstratos e formais (concordância nominal e verbal, morfologia derivacional, pronome de caso e formas de gênero, a maioria das preposições) e a presença de elementos que são reconstruídos por meio de gramaticalização (artigos, formas generalizantes de tempo, aspecto e modo, pronomes interrogativos, pluralizador, pronomes para todas as pessoas e números, dentre outros).

A proposta defendida por Bickerton teve bastante relevância para a crioulística na década de 1980 porque abrangeu características gerais dos crioulos e, segundo o próprio Bickerton, proporcionou aos estudos dessas línguas a possibilidade de associação às teorias da aquisição, uma vez que deu visibilidade às situações de atuação da GU na transmissão geracional da linguagem.

No entanto, a hipótese do Bioprograma recebeu diversas críticas devido, principalmente, à forma como Derek Bickerton classificou as línguas *pidgins* e crioulas como línguas simples e empobrecidas gramaticalmente apenas por não apresentarem as características gramaticais da LA. O preconceito linguístico se estende aos falantes dessas línguas ao serem chamados, pelo autor, de maneira enfática como não-

européus, levando a crer que somente os europeus (os dominadores) eram detentores de poder, e todos os outros povos, africanos em sua maioria, seriam inferiores por não serem europeus.

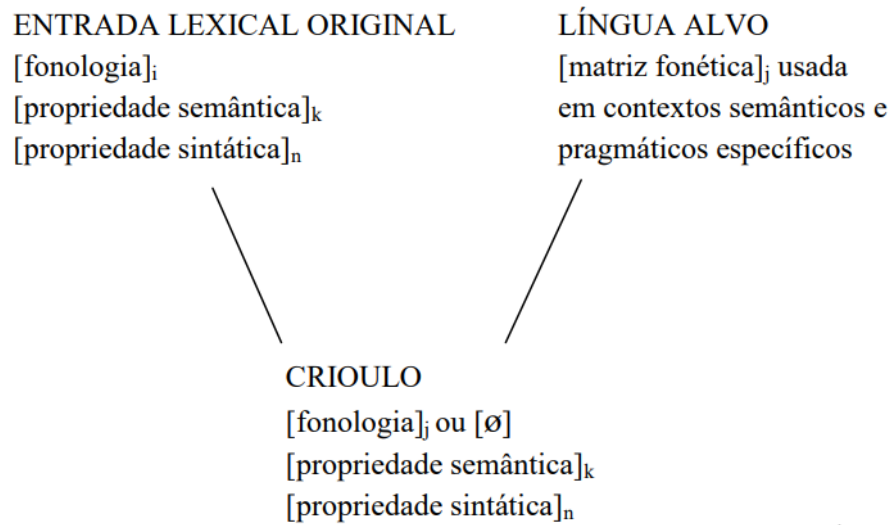
Além dessa crítica, muitos foram os questionamentos acerca de que as crianças seriam os agentes fundamentais da formação do crioulo. Os estudos de cunho substrativista (LEFEBVRE, 1998, 2001; LEFEBVRE E LUMSDEN, 1994), por exemplo, discordam desse aspecto e propõem que são os adultos falantes de línguas do substrato que desempenham o papel principal na formação de variedades linguísticas que se formam em situações de contato intenso, radical e abrupto, transferindo os mecanismos gramaticais de sua(s) língua(s) nativas, no processo de pidginização/crioulização.

### 3.1.2 Relexificação

Considerando o substrato como aspecto fundamental na formação de *pidgins* e crioulos, Claire Lefebvre dedicou-se a estudar sistematicamente por mais de 20 anos o crioulo haitiano, que tem como língua lexificadora o francês e como principal língua de substrato o fongbe, língua nativa do oeste africano. Em sua pesquisa, Lefebvre respaldou-se na hipótese da relexificação, que é um processo mental através do qual é construído um novo léxico da seguinte maneira: as formas fonológicas das entradas lexicais de um crioulo provêm de expressões do superstrato, enquanto as propriedades semânticas e sintáticas dessas entradas lexicais seguem o padrão das línguas de substrato (MUYSKEN, 1981, p. 62 *apud* Lefebvre 1998, p. 16).

Lefebvre (1997, 1998, 2001) segue a ideia proposta por Muysken (1981) e busca explicar o papel do substrato na gênese crioula. A autora defende, com base nos dados haitianos que envolvem *inputs* lexicais de categoria funcional, que o processo mental de relexificação responde por essa divisão de propriedades de maneira direta. É importante ressaltar que a relexificação é direcionada pela semântica, logo, para que ela ocorra, é necessário que os *inputs* lexicais da L1 e da L2 tenham algo em comum semanticamente, como mostra o esquema apresentado por Lefebvre (2001), evidenciado em (1) e o exemplo do crioulo haitiano em (2).

(1)



(LEFEVBRE, 2001, pág. 12)

Primeiro, na fase inicial do crioulo, o processo de relexificação prevê um léxico comum para que os falantes das línguas de substrato possam se comunicar. Os falantes, criadores do léxico crioulo, identificam, na LA, uma sequência fonética apropriada que sirva como uma representação fonológica para a entrada lexical copiada da(s) língua(s) nativa(s).

(2)

HAITIANO	FRANCÊS	FONGBE
<i>vyann</i>	<i>viande</i>	<i>làn</i>
‘carne’	‘carne’	‘carne’
‘animais comestíveis’		‘animais comestíveis’
(complemento do verbo ‘matar’)		(complemento do verbo ‘matar’)

(LEFEVBRE, 2008, pág. 199)

No exemplo (2), as entradas lexicais do francês e do fongbe, compartilham o significado de *carne*, sendo o novo item lexical do crioulo haitiano derivado da fonologia do francês, mas com propriedades sintáticas e semânticas da língua de substrato, o fongbe. Ainda de acordo com a hipótese relexificação, os aprendentes 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).

No entanto, por sua própria natureza, a relexificação não pode ser o único processo envolvido na gênese crioula, mesmo no caso de um crioulo radical como o haitiano. Lefebvre (2001) afirma que a relexificação se aplica na gênese crioula quando os falantes das línguas de substrato têm como alvo a língua de superstrato, e quando esses falantes começam a mirar nos léxicos relexificados, isto é, no crioulo inicial, não estão mais usando a relexificação para desenvolver o crioulo. Por esse motivo, a autora propõe que, quando os falantes das línguas de substrato começam a direcionar-se para os léxicos relexificados, dois outros processos desempenharão um papel no desenvolvimento do crioulo: o nivelamento de dialetos e a reanálise.

O processo de nivelamento ocorre durante o contato entre línguas. Devido as diferenças linguísticas do substrato de uma língua crioula, o léxico criado por relexificação, no início da formação do crioulo, não é uniforme. Assim, o nivelamento dialetal atua reduzindo a variação resultante da relexificação dos vários léxicos do substrato. A reanálise é um processo através do qual uma forma fonológica específica associada a uma entrada lexical passa a ser associada a outra entrada lexical - esse processo também é chamado de gramaticalização. Apesar de não discutir sobre o papel dos fatores sociais, a autora afirma que eles interagem com os processos envolvidos na formação dos crioulos, principalmente na fase do nivelamento dialetal.

Vale salientar que Lefebvre (1998) considera que o falante adulto é o principal agente da formação das línguas crioulas, mas pondera que, assim como em outras situações de mudança linguística, as crianças têm um papel importante no desenvolvimento de crioulos.

A hipótese da relexificação recebe algumas críticas por se tratar de um processo mentalista que não se aprofunda nas questões sócio-históricas. DeGraff (2002) diz que Lefebvre falha ao não levar em conta de forma mais evidente o papel da sócio-história na descrição do processo de relexificação no crioulo haitiano, sendo uma hipótese excepcionalista. Segundo DeGraff, deve-se considerar a aquisição de uma L2 com acesso a um *input* restrito analisando os graus variados que podem existir

na relação substrato L1/superstrato L2, justamente porque a proposta da relexificação evidencia o papel de falantes adultos.

Sobre esse aspecto, DeGraff (2002) salienta que não se pode ignorar o papel sociolinguístico das crianças na gênese do crioulo, pois a tentativa de aquisição dos falantes adultos juntamente com as inovações linguísticas das crianças podem gerar uma rede de novos itens gramaticais na ecologia linguística<sup>9</sup> da nova língua, que não são encontradas no substrato nem no superstrato. Outra crítica feita por DeGraff (2002) é sobre a concepção de que antes da formação de um crioulo, os falantes utilizam uma interlíngua<sup>10</sup> sujeita às restrições da GU.

É importante ressaltar que DeGraff (2002) reconhece a importância da hipótese da relexificação por considerar a influência do substrato na formação do crioulo porque não se pode negar as influências dos substratos africanos, entretanto, explica que é preciso discutir sobre as interações na aquisição (precoce) de L2 e L1 e as implicações dessas interações no processo de formação de línguas crioulas.

### 3.1.3 Transmissão Linguística Irregular

Baxter (1995), e Baxter e Lucchesi (1997), postularam o termo *transmissão linguística irregular* (TLI) para designar a integração da transferência do substrato e o processo de nativização na gênese das línguas crioulas, considerando processos históricos de contato massivo entre línguas de tipologias diferentes (BAXTER, 1992, 1995; BAXTER E LUCCHESI, 1997; LUCCHESI, 2003; 2009). Essencialmente, a TLI postula que os elementos gramaticais podem surgir no processo de transferência do substrato, mas, no plano da fala dos adultos, mantêm-se em uma situação instável e de muita variação, e só são efetivamente gramaticalizados quando integrados no processo de aquisição da língua materna das crianças que nascem na situação do contato.

---

<sup>9</sup> Haugen (1972) propôs a ecologia da língua ou ecologia linguística como o estudo da interação de qualquer língua a seu meio ambiente.

<sup>10</sup> De acordo com Schütz (2018), interlíngua é o sistema de transição criado pelo aprendiz, ao longo de seu processo de assimilação de uma língua estrangeira. É a linguagem produzida a partir do início do aprendizado, caracterizada pela interferência da língua materna, até o aprendiz ter alcançado seu teto na língua estrangeira, ou seja, seu potencial máximo de aprendizado. A ocorrência de interlíngua é significativamente maior em adultos do que em crianças.

Na TLI, a geração de falantes mais velhos fornece dados linguísticos primários (DLPs) a partir do seu desempenho na língua nativa, ou seja, os parâmetros da gramática das crianças são baseados no modelo da gramática dos pais. Porém, no processo de crioulização, os DLPs para a aquisição da língua materna são fornecidos por adultos que adquiriram a L2 através do contato linguístico intenso (em situação de aprendizagem precária e/ou forçada, no qual a transmissão geracional ocorre em condições adversas de aprendizado), sendo que essa L2 contém material linguístico distinto daquele da LA. 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 suposição dos autores é que os dados linguísticos dos aprendentes adultos mostram um acesso parcial à GU, levando-os a preencherem as lacunas da L2 com parâmetros da língua materna. Isso faz com que ocorra um *continuum* para a formação de uma nova variedade linguística, pois as gerações de futuros falantes aprendem uma nova língua, originada de um *input* escasso, com a possibilidade de preencher as lacunas pelo processo de reanálise (LEFEBVRE, 2001).

Para Lucchesi e Baxter (2009), nos casos de contato intenso entre línguas podem ocorrer situações que levam à formação de um crioulo radical, ou não, pois é possível haver um processo de TLI do tipo mais leve, como os autores sugerem ter ocorrido na variedade do PB, em que não se originou outra identidade linguística, mas uma variedade da língua alvo, isto é,

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 (LUCCHESI e BAXTER, 2009, p. 102).

Segundo Lucchesi e Baxter (2009), a TLI pode ser percebida quando há uma simplificação gramatical (BICKERTON, 1998) e uma redução na estrutura gramatical da língua devido 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).

Na hipótese da TLI, as estruturas mais abstratas que não possuem funcionalidade no ato da comunicação são dispensados, dessa maneira, as marcações morfológicas de caso, número, pessoa, tempo, modo e aspecto, e as relações sintáticas de regência, concordância, coordenação e subordinação são dispensados. Em suma, se formaram variedades de L2 com particularidades gramaticais distintas da LA, dado o fato do contexto sócio-histórico em que originaram.

Alguns aspectos sobre a TLI podem ser observados, como a proposta de uma simplificação gramatical da variedade linguística formada em situação de contato, crítica atribuída também à hipótese do Bioprograma da Linguagem (BICKERTON, 1988), explicada na subseção 3.1.1, bem como a defesa de que os crioulos são originados através de uma aquisição defectiva. Isso é questionável, pois, segundo Mufwene (2002) e Aboh e Ansaldo (2017), não há aquisição imperfeita de línguas, mas sim contextos sócio-históricos e sociolinguísticos distintos.

### 3.1.4 Competição e Seleção – Ecolinguística

Uma noção discutida na crioulística é que as línguas crioulas surgem a partir da nativização<sup>11</sup>, porém hipóteses mais recentes têm sugerido que os crioulos são línguas desenvolvidas a partir dos mesmos processos de formação de línguas não crioulas. Uma hipótese que defende essa proposição é sugerida por Mufwene (2001, 2002), a qual propõe que todas as línguas, incluindo crioulos, são formadas a partir dos processos de *Competição e Seleção*, considerando a *Evolução Linguística*.

Mufwene (2002) propõe que a competição e a seleção de traços, bem como o hibridismo<sup>12</sup>, desempenham papéis importantes na evolução linguística. O autor afirma que o hibridismo é uma condição normal em todas as línguas, não somente nos crioulos, como dizem alguns estudos. A diferença está no fato de que as línguas crioulas apresentam hibridismos mais evidentes do que as línguas que não têm

---

<sup>11</sup> Autores como Mufwene e Aboh tem se dedicado ao estudo das línguas crioulas a partir de uma visão não excepcionalista, defendendo que o processo de nativização dessas línguas ocorrera de forma semelhantes a outras línguas naturais não africanas, com a diferença do contexto sócio-histórico.

<sup>12</sup> Processo de combinação de variedades de línguas tipologicamente distintas, resultante da competição e seleção de traços linguísticos (MUFWENE, 2002).

contato linguístico intenso. Mufwene considera que o hibridismo é um processo natural e que envolve a seleção de traços. O autor utiliza o termo *feature pool* para explicar que os falantes teriam à disposição uma espécie de piscina de recursos ou traços formada a partir da combinação de características das línguas em situação de contato.

Segundo o autor, nessa piscina de traços, pode haver congruência quando as línguas possuem características tipológicas semelhantes, porém quando a tipologia é diferente, há uma competição de traços de maneira que o domínio de uma língua sobre a(s) outra(s) é mais evidente. A congruência, que favorece características comuns entre as línguas, é, em si, uma forma de seleção, pois algumas variantes não congruentes são mais propensas a serem descartadas, principalmente quando as populações que não falam a língua alvo são socialmente integradas.

Mufwene (2002) destaca que, havendo ou não congruência, toda nova variedade colonial (crioula ou não) derivada de uma língua europeia é fundamentada pelo contato e pelos mesmos mecanismos de evolução linguística e seleção de traços. O autor diz que o processo de seleção é possível devido à própria natureza da língua enquanto sistema complexo que se adapta a partir de estruturas multimodulares. Uma questão central dos estudos sobre crioulos, segundo o autor, gira em torno de quais seriam os princípios ou restrições aplicadas de forma inconsciente pelos falantes e sob quais condições particulares as seleções individuais se tornam seleções coletivas e, posteriormente, normas na comunidade de fala.

O autor critica os termos “transmissão” e “aquisição” nos estudos sobre mudança linguística, pois, para ele, esses termos são imprecisos e não mensuram todo o processo de evolução linguística. Mufwene prefere utilizar a expressão “herança de genes, de características linguísticas”, fazendo analogia a princípios da Biologia e da Ecologia, herança no sentido de mudança. Nessa perspectiva, o autor compara a evolução linguística com a reprodução animal, mostrando que, nesta a prole não é ativa no processo, mas naquela as características dos genes são ativas e se comportam de maneira seletiva, criando e recriando possibilidades mediante à replicação diferente das unidades do sistema.

Portanto, a evolução das línguas é um processo criativo. Mufwene (2002) ressalta que o desenvolvimento das línguas crioulas é normal e natural e rejeita a ideia de Thomason e Kaufman (1988) de que os crioulos não podem receber classificação genética porque são línguas formadas por lexificação, bem como a proposição de Bickerton (1988) de um bioprograma da linguagem que considera o crioulo como uma

língua defectiva que se origina como tentativa de segunda língua por falantes adultos e se torna a língua materna das crianças nascidas posteriormente. Mufwene (2002) concorda, em certa medida, com as abordagens que consideram o papel da língua alvo como base lexificadora dos crioulos com a influência gramatical das línguas de substrato, porém, o autor diz que durante o processo de aquisição, a herança lexificadora é modificada pela influência do substrato ou do próprio sistema linguístico emergente.

Mufwene explica que, no processo de aquisição, os falantes vão fazendo combinações possíveis e, naturalmente, vão percebendo restrições: em determinada língua um traço é possível, em outra língua, não. Conforme explica o autor, a competição e a seleção de traços é o que faz surgir uma nova língua ou promove a mudança linguística dentro de uma mesma língua, salientando que a diferença na formação do crioulo para outras línguas é somente no que concerne às condições sob as quais os falantes adquirem a base lexificadora, sendo o processo de evolução sistêmica o mesmo para todas e quaisquer línguas.

Além da crítica aos estudos linguísticos que não consideram os crioulos nas classificações genéticas, Mufwene (2002) questiona o porquê de os linguistas não denominarem como crioulas as línguas românicas, por exemplo, visto que elas também foram formadas a partir de situação de intenso contato e dominação europeia, como se o termo “crioulo” fosse apenas destinado às línguas que possuem substratos africanos. Para explicar o desenvolvimento linguístico de um crioulo pelo processo de seleção, o autor, primeiramente, afirma que esse desenvolvimento se inicia pela “escolha” da língua alvo, dentre aquelas que estão em contato.

Isso pode ser observado tanto na aquisição da língua pelas crianças como na aquisição de segunda língua pelos adultos, levando em conta o caráter gradual do processo, com a incrementação do léxico e dos diferentes componentes da gramática emergente. De acordo com essa visão, quando um indivíduo adquire uma língua diferente da sua língua nativa, cada unidade do sistema entra em concorrência, o que, em larga escala, pode gerar mudança no idioleto e também no dialeto da comunidade.

Mufwene (2002) argumenta que a língua é adquirida individual e coletivamente, sendo que os idioletos carregam características naturais dos falantes que podem resignificá-las em diferentes momentos da vida, disso decorre o porquê de ele considerar aspectos da ecologia para a mudança linguística. Embora o autor afirme e reafirme que a evolução das línguas ocorre por meio de seleção de traços, ele ainda

questiona: qual a dinâmica das seleções individuais e como elas passam a ser coletivas de maneira que determinam a trajetória evolutiva de uma língua?

Para tentar responder a essa questão, o autor recorre ao processo de hibridismo das línguas, utilizando como exemplo os crioulos porque neles é possível verificar de forma mais evidente como a influência de línguas distintas pode produzir uma nova variedade de língua semelhante ou não, observando a congruência e a reestruturação guiada por seleção. Outra proposição do autor é observar as restrições que ocorrem dentro do processo de seleção a partir da combinação de traços linguísticos sob condições específicas de competição.

A restrição de um determinado aspecto gramatical pode ocorrer pelo fato de os falantes perceberem naturalmente qual a opção mais viável dentro da estrutura linguística, qual a opção mais transparente, ou regular, ou mais saliente, ou não semanticamente vazia. Uma ressalva feita pelo autor é que as restrições não ocorrem de maneira uniformizada, pois existem crioulos formados com a mesma base lexicadora, mas que são diferentes porque não possuem a mesma seleção de variantes. O objetivo do autor é mostrar que a competição e a seleção são inerentes à evolução das línguas, ocorrendo no indivíduo e na coletividade.

Apesar de ser uma hipótese muito perspicaz e valorizada na crioulistica recente, a ideia de competição e seleção de traços sofre algumas críticas, a exemplo do que mencionou McWhorter (2014). Segundo ele, a ideia da piscina de traços não explica tudo, pois muitas características linguísticas que são encontradas em línguas crioulas que não são oriundas do substrato nem do superstrato. Dessa forma, a questão colocada por McWhorter é que se a gênese crioula ocorre a partir de uma piscina de traços, como seria explicada a morfologia simplificada de certas línguas crioulas, como o Palenquero<sup>13</sup>, que não apresenta necessariamente traços do Kikongo (língua de substrato) nem do Espanhol (língua de superstrato).

McWhorter (2014) critica a hipótese proposta por Mufwene, pois em estudos anteriores (1998, 2001, 2005), McWhorter sugere um paradigma chamado de Protótipo Crioulo, em que o autor descreve os crioulos como línguas de gramáticas simples, desprovidos de aspectos de complexidade e que, supostamente, podem surgir com o tempo. Para McWhorter (2014), as línguas mais antigas possuem morfologia, enquanto que os crioulos não tiveram tempo suficiente para adquirir essa

---

<sup>13</sup>Palenquero ou palenque é uma língua crioula de base espanhola e portuguesa, falada na Colômbia, sendo a única língua crioula falada na América Hispânica.

complexidade porque são línguas novas. Muitas são as críticas a respeito desse posicionamento, principalmente porque McWhorter não explica qual o conceito de “simplicidade” e de “complexidade” abordados por ele quando se fala em gramática, em morfologia.

Respondendo à crítica feita por McWhorter, Mufwene (2014) diz que a hipótese da piscina de traços não é um mecanismo, mas sim um estado de coisas que surge do contato entre línguas na mente dos falantes, os quais têm acesso aos recursos mais relevantes dos sistemas linguísticos em contato. Mufwene acredita que a combinação e a seleção de traços podem não dar conta de explicar tudo sobre o surgimento dos crioulos, pois se trata de um processo complexo e prolongado, impulsionado por vários fatores interativos. Mufwene discorda de McWhorter a respeito da simplificação gramatical como explicação da gênese crioula, visto que, além da (re)combinação de traços, existem os processos de hibridização, gramaticalização, contração, enfraquecimento, reforço etc. (processos ignorados por McWhorter).

### 3.1.5 Teoria Nula

Na tentativa de desconstruir a ideia de excepcionalidade atribuída às línguas crioulas, os trabalhos de DeGraff (2003), Aboh e Ansaldo (2007), Aboh e Matthews (2007), Aboh (2015) e Aboh e DeGraff (2017) seguem a proposição de que princípios linguísticos gerais, linguística histórica e mudança linguística são perspectivas necessárias para observar e explicar o surgimento de uma língua em situação de contato, principalmente quando o contato é intenso. Esses autores concordam com o que propôs Mufwene e também consideram a perspectiva da piscina de traços e discordam da simplificação gramatical abordada em muitos estudos crioulistas.

O objetivo principal dos estudos desses autores é superar o que eles chamam de dicotomia artificial entre línguas crioulas e não crioulas. Para isso, Ansaldo e Matthews (2007) fazem algumas considerações críticas a respeito dos estudos crioulistas que tendem a ser diferentes dos linguistas que trabalham em outros campos de estudos, destacando que enquanto os não-crioulistas normalmente trabalham com a suposição de que seus trabalhos contribuirão de alguma forma para o campo da linguística como um todo, os estudos sobre línguas *pidgins* e crioulas

parecem promover a manutenção de uma crença que os objetos de seus estudos são únicos. Essa visão coloca as línguas *pidgins* e crioulas como algo exótico e as isola, fato que, segundo os autores, contribui para uma perspectiva distorcida sobre formação e aquisição dessas línguas.

Segundo Aboh e Ansaldo (2017), o principal mito em torno da gramática das línguas crioulas é a ideia de simplificação. Os autores afirmam que é característica comum dos crioulos não possuir flexões se comparados às línguas lexificadoras, sendo a ausência de morfologia flexional frequentemente considerada como uma evidência típica dos crioulos. Muitos estudos abordam o processo de simplificação relacionado com uma aquisição imperfeita de L2 devido ao acesso limitado à gramática da LA.

Criticando essa ideia, Aboh e Ansaldo (2007) destacam duas falhas: i) a maioria dos estudos que consideram essa perspectiva concentram-se no paradigma verbal, sendo que um olhar apurado para o paradigma nominal sugere uma outra perspectiva; ii) o papel da congruência em favorecer ou desfavorecer a morfologia flexional nos crioulos nunca foi estudado.

Por exemplo, línguas de substrato do grupo kwa não possuem morfologia flexional diferente das línguas lexificadoras (francês e inglês). A falta de flexão nos crioulos resultantes desse contato pode ser explicada pela possibilidade da seleção de traços feita pelos falantes. Isso não significa uma aquisição imperfeita. Para exemplificar a proposta da piscina de traços, os autores utilizam os crioulos do Suriname e do Sri Lanka e seus respectivos sistemas nominais, argumentando que a falta de certas flexões nominais nos crioulos não é resultado de uma mera simplificação, mas sim consequência do processo de concorrência e seleção de traços sintáticos e semânticos das línguas em contato.

Os autores propõem que as características semânticas são as que mais entram em competição, como tópico, foco, especificidade, modalidade etc. Traços mais sintáticos e funcionais, puramente formais, como pessoa e número, tendem a ser analisados como menos concorrentes, logo são desfavorecidos em situação de contato. Isso ocorre porque os traços de função semântica são mais proeminentes no discurso do que os traços puramente formais. Aboh e Ansaldo (2007) destacam que as características dos crioulos ou de outras línguas formadas em situação de contato menos evidente são combinações e recombinações de traços linguísticos que emergem do processo de reestruturação.

Os autores citam dois casos distintos de formação de crioulo para comprovar a competição e a seleção de traços: i) o caso dos crioulos do Suriname: a tipologia das línguas lexificadoras (inglês e português) possui semelhanças com as línguas nativas do grupo *kwa*, como a ordem de palavras e a ausência de morfologia flexional em determinados contextos; isso pode ser verificado também nos crioulos que surgiram desse contato; ii) o caso do agrupamento de dois sistemas diferentes: como a marcação de caso no cingalês e no tâmil do Sri Lanka, que é resultado de uma mudança tipológica radical do malaio para a gramática do lankan.

Aboh e Ansaldo (2007) criticam a abordagem dos estudos linguísticos que consideram todos os crioulos como línguas que possuem estruturas semelhantes independentemente do processo de reestruturação. Assumindo uma perspectiva diferente, eles argumentam que as características tipológicas dos crioulos diferem na estrutura justamente porque o processo de seleção de traços pode ser diferente em cada situação de contato.

Considerando a visão Croft (2000), os autores reforçam a ideia de que as línguas são um conjunto de idioletos e, quando os falantes interagem, há um compartilhamento de características linguísticas, as quais podem ser replicadas ou não com alterações, inovações, dando o caráter evolutivo da linguagem. A replicação pode ser quase que idêntica entre os falantes quando o ambiente tem um alto grau de monolingüismo e uma alta tendência normativa, onde o contato é menos saliente. Já em populações mais migratórias, o ambiente de multilingüismo favorece replicações mais profundas, mais alteradas devido ao contato intenso.

Para entender o que determina a direção dos processos de replicação profunda, os autores recorrem à noção de piscina de traços para representar a mistura de códigos linguísticos aos quais os falantes de um ambiente multilíngue podem ser expostos, levando em conta também a mistura de códigos nos idioletos. São esses dois níveis de exposição de traços estabelecidos pelos autores: individual e coletivo. Para fins metodológicos, Aboh e Ansaldo observaram a mistura de traços apenas no nível coletivo, discutindo a questão da tipologia linguística para tentar avaliar o processo de seleção de traços. Segundo eles, os traços mais marcados são mais competitivos e tendem a se reforçar e a se tornar regulares na língua emergente.

Rejeitando a ideia de simplificação morfológica dos crioulos como algo imperfeito, os autores afirmam que essas línguas são resultado de um processo de competição, seleção e combinação de traços bem-sucedidos, visto que morfemas

semanticamente ativos são favorecidos e os semanticamente vazios tendem a ser descartados, ou seja, não fazem “falta” na língua emergente

Para exemplificar essa assertiva, os autores elencam algumas características gramaticais dos crioulos de Suriname, do inglês (língua alvo) e do gungbe (língua de substrato): o inglês e o gungbe se distinguem quanto à marcação de número. Em gungbe, o número é representado apenas por um marcador e o substantivo permanece em sua forma simples; em inglês, o número é sempre marcado no substantivo (cf. ABOH e DEGRAFF, 2017, p 55). Os crioulos resultantes do contato entre essas línguas não apresentam morfologia flexional em decorrência da falta de competição de traços de número nas línguas originárias. Dessa forma entende-se que o crioulo não é resultado de imperfeição, mas sim resultado da combinação de traços semanticamente relevantes.

Outrossim, é que, em situação de contato, a sintaxe e a semântica de categorias funcionais são desmontadas e remontadas de várias formas que, não necessariamente, correspondem às combinações encontradas nas línguas de origem, como é o caso de algumas propriedades gramaticais dos crioulos do Sri Lanka, as quais não são típicas da língua alvo nem da língua de substrato, segundo os autores. Eles consideram ainda que a função das categorias funcionais está sujeita a restrições diferentes em uma situação de competição, reafirmando que a formação dos crioulos não deriva meramente de aquisição/perda/aquisição/reestruturação, e sim de recombinação geral de características linguísticas.

Do ponto de vista mais sócio-histórico, Ansaldo e Matthews (2007) dizem que as teorias crioulas foram ordenadas pelo "olhar normativo" da Europa em relação ao mundo não-branco e não-cristão, inclusive citam um estudo de 1894, feito por Poyen-Bellisle, na Universidade de Chicago, que descreve como os escravizados tentam imitar o discurso do homem civilizado, mas falham porque eles não pertencem à mesma raça e possuem características físicas distintas, como os lábios grandes. Nota-se um preconceito racial evidente nesse estudo, o que, infelizmente, perdurou por muito tempo e caracterizou negativamente estudos sobre os povos africanos, bem como suas línguas nativas e as línguas crioulas como línguas menos importantes. Para DeGraff (2003), apesar de absurdos, alguns mitos em torno das línguas crioulas persistem nos estudos linguísticos, por isso os crioulos são considerados "órfãos" não genéticos fora da família árvore das línguas humanas. Alguns estudos recentes



(MUFWENE, 2001, 2002, 2008) têm buscado desconstruir essas ideias, atribuindo aos crioulos à genealogia linguística.

Diante das considerações feitas para desconstruir os mitos em torno da formação das línguas crioulas, Aboh e DeGraff (2017) denominam de Teoria Nula a hipótese que apresentam justamente por afirmar que não há nenhuma estipulação *sui-generis* para explicar a origem dos crioulos, pois todas são assunções aplicáveis a qualquer língua. Portanto, nessa abordagem, a emergência de qualquer nova língua ou variedade no contexto de contato linguístico chama atenção para os contextos sócio-históricos de formação das línguas e para o papel da aquisição de L1 e de L2 na construção de novas gramáticas a partir de *inputs* variáveis e complexos.

Segundo esses autores, a única diferença ideologicamente plausível e empiricamente fundamentada entre as línguas crioula e não crioula é a sócio-história. No demais, quanto às questões linguísticas, falantes individuais em ecologias de contato, seja em línguas crioulas, romance ou germânico, utilizariam o mesmo processo mental na formação de novas línguas.

### **3.1.6 Uma outra proposição aquisicional**

Com o objetivo de discutir a aquisição de L2 e a formação de *pidgins* e crioulos e buscando contribuir para uma melhor compreensão dos processos de L2 associados às características de interlínguas, Baptiste (2016) confronta estudos já realizados sobre o contato linguístico e a formação de crioulos e tenta propor uma junção de teorias para propor um novo olhar sobre os processos e os rótulos que são atribuídos à aquisição e à origem dos crioulos. Em outras palavras, ao reavaliar estudos mais antigos e estudos mais recentes, Baptiste (2016) espera apresentar um equilíbrio entre as visões a respeito da origem, da formação e da aquisição dos crioulos.

A autora destaca a perspectiva de Schuchardt na investigação da origem e da estrutura de *pidgins* e crioulos considerando a contribuição dos superstratos e dos substratos e o papel dos universais linguísticos na aquisição e desenvolvimento dessas línguas. A proposta de Schuchardt é que o processo de simplificação atinge todos os níveis da gramática, exemplificado da seguinte maneira: se a L1 tem regras Y e a L2 tem Y+1, o resultado de qualquer processo de pidginização favorece a forma da L1. Tal processo implica em dizer que não somente os recursos da L1 são

transferidos para a L2, mas também que os aprendentes tendem a optar pela forma mais simples, com menos regras. Para Baptiste (2016), a ideia de transferência da L1 para a L2 e processos de aquisição de L2 ligados à simplificação gramatical é tão antiga quanto à linguística de contato, embora pareça estar ancorado como algo contemporâneo.

A autora cita o trabalho de Adolfo Coelho (1880) que se baseia em um aspecto da perspectiva de Schuchardt: a visão genérica de que os crioulos se desenvolvem a partir dos mesmos processos, independentemente das línguas das quais se originaram. Baptiste (2016) critica essa visão, pois o conceito de simplicidade gramatical não é profícuo para os estudos linguísticos.

Sobre os aspectos da aquisição de L2 envolvidos na formação de línguas crioulas, Baptiste (2016) diz que é possível identificar três pontos contraditórios i) a aquisição de L2 como aquisição do léxico da língua alvo (CHAUDENSON, 2001); ii) os falantes do crioulo inicial adquirem uma L2 que não é necessariamente a língua alvo (MUYSKEN, 2001); e iii) os crioulos seriam resultados da capacidade de tentativas de aquisição de L2 e representariam estágios iniciais da interlíngua (PLAG, 2008).

Diferentemente dessas três perspectivas, alguns linguistas acreditam que os crioulos emergem das interlínguas e não envolvem processos de transferência (MEISEL, 1983). Outros propõem que adquirir o léxico era a menor das prioridades dos falantes, o objetivo era criar um meio de comunicação interétnico (BAKER, 1995). Baptiste, então, questiona: que língua seria a base de formação dos crioulos? Qual seria a L2 alvo? Por que há menção de aquisição de língua "imperfeita" ou "incompleta" com todas as conotações infelizes que tais expressões transmitem quando há poucas evidências? Que língua de superstrato era acessível? E, mesmo que fosse possível alcançá-la, fazia parte das intenções e prioridades originais dos interlocutores? Se havia uma língua alvo e essa língua não era de superstrato, o que era então? Diante de tantas questões, muitos estudiosos continuam incertos quanto à natureza do L2.

Na tentativa de responder a essas questões, Baptiste (2016) faz um resumo das hipóteses superstrativistas e das substrativistas, agregando trabalhos sobre aquisição de L2 na perspectiva de processos psicolinguísticos. A proposição mais criticada por Baptiste (2016) refere-se ao que propuseram Plag (2008) e Pienemann *et al* (2005) ao tratarem os crioulos como interlínguas. Para confrontar essa hipótese,

Baptiste (2016) analisa os fenômenos da ordem de palavras e da negação em algumas línguas crioulas e como estão representadas no Atlas de estruturas linguísticas de *pidgins* e crioulos, APiCS (MICHAELIS *et al.* 2013).

A análise desses fenômenos feita por Baptiste (2016) evidencia que não se pode alegar que existe um protótipo crioulo e, portanto, especular que o mesmo conjunto de processos e mecanismos deu origem uniforme a todos eles. Além disso, os *pidgins* e crioulos não exibem características típicas de interlínguas, então há pouco objetivo em retratá-los como um estágio da língua em desenvolvimento, como se estivesse preso entre dois mundos linguísticos, pois haveria uma L1 abandonada e uma L2 não conquistada. Baptiste (2016) diz que considerar os crioulos como interlínguas, como se resultassem de aquisição de segunda língua 'interrompida', 'incompleta', 'imperfeita', não faz jus à rica diversidade de seus recursos no APiCS.

É importante ressaltar que, para a aquisição de L2, o objetivo é adquirir a segunda língua, enquanto que no uso da segunda língua, a preocupação é com a forma como os falantes utilizam seu conhecimento vigente de uma L2 juntamente com outros tipos de conhecimento (incluindo a L1) para conceber uma nova maneira de comunicar naquela língua. No caso do uso de L2, o principal objetivo dos falantes é comunicar e não necessariamente adquirir uma segunda língua em si.

Após ponderar os principais aspectos de vários estudos, Baptiste diz que as pesquisas de Schuchardt (1883; 1909; 1914), Baker (1994), Aboh (2009), Aboh e DeGraff (2014) têm em comum a possibilidade de associação para uma abordagem substratista e de criatividade à gênese crioula, considerando que é necessário retomar a ecologia linguística de cada crioulo (MUFWENE, 2001) como um ponto de partida, além de levar em conta as línguas de origem em contato, suas características, a demografia da população, suas relações de poder e a atitude dos falantes (THOMASON, 2001).

Baptiste (2016) enfatiza que se os processos de aquisição de L2 não forem considerados nos estudos sobre *pidgins* e crioulos, a comparação dessas línguas com a LA pode ser irrelevante, pois continuariam sendo disseminadas ideias de que línguas *pidgins* e crioulas são versões "simplificadas" de seus superstratos, sendo esse um rótulo que precisa ser retirado da formação dessas línguas, as quais são línguas naturais em seu direito próprio.

Elencadas as principais teorias crioulisticas acerca do contato linguístico, sistematizo, na seção 3.2 a seguir, os principais fatos sócio-históricos sobre o

povoamento e formação do Brasil, principalmente, da Bahia para elucidar acerca dos efeitos do contato linguístico na formação do PB.

### 3.2 PRINCIPAIS ASPECTOS SÓCIO-HISTÓRICOS DA FORMAÇÃO DO BRASIL

As tomadas de territórios por Portugal, por volta do século XV, não tinham finalidades de povoamento, mas sim de expansão do seu império mercantil. Segundo Ferro (2008), o que os portugueses queriam era o império do comércio marítimo e não necessariamente as terras. Por isso, estabeleciam pequenos enclaves, os quais tinham a função de servir de entreposto para a movimentação do comércio entre as populações locais, bem como servir de apoio às frotas comerciais.

No início do século XVI, aquilo que os portugueses dispunham no Índico era um conjunto muito diversificado e descontínuo de centros urbanos, fortalezas, feitorias e outras formas de domínio que visavam assegurar o controle da “navegação” e “comércio” cujo “senhorio” os reis portugueses reivindicavam como título seu por direito de conquista. E que mantinham, em larga medida, pela força, ou seja, por ameaça militar constante. Este conjunto de pontos nevrálgicos, com vários estatutos, num imenso espaço geográfico, tem sido comparado pelos historiadores a uma “rede” ou “nebulosa”, aliás com muitos antecedentes na região. Pouco tinham que ver, portanto, com o tipo de domínio territorial alargado, como era comum na Europa e também nos espaços tutelados pelas monarquias nas Américas. (MONTEIRO, 2010, p. 218-219)

A supremacia naval portuguesa e a estrutura em rede fizeram com que um país relativamente pequeno conseguisse dominar por quase um século o comércio marítimo em territórios que se estendiam da costa africana à Índia, Malaca, Timor-Leste e Macau. Porém, no século XVI, os holandeses expandiram sua rota nesse percurso em declínio do domínio português. Dessa forma, o uso da língua portuguesa em território asiático foi reduzido.

Conforme o domínio dos holandeses se estendia pela Ásia, Portugal perdia espaço no comércio das especiarias e começava a mudar as características das navegações comerciais para navegações com fins de ocupação de territórios, estabelecendo presença no Atlântico Sul. De acordo com Alencastro (2000), o domínio português estabelecido nessa região se deu em dois eixos: (i) a ocupação do Brasil, com uma espécie de economia de coleta, com o corte do pau-brasil e a

exploração do trabalho indígena; (ii) o tráfico de africanos escravizados que sustentava a intensa produção açucareira.

Logo, no Brasil, Portugal não se firmou apenas por comércio, mas estabeleceu uma forte colônia de exploração e povoamento. Ancorado pelos entrepostos da costa africana, comerciantes portugueses e africanos praticamente não tinham concorrência em relação ao tráfico de africanos. Disso decorre a intensidade desse marcante fato sócio-histórico em relação à formação do Brasil.

A seguir, darei ênfase a esses dois processos importantes para a formação do Brasil, e do Estado da Bahia: a exploração e o genocídio das nações indígenas e o tráfico de africanos escravizados. A sócio-história do Brasil é complexa, envolve muitos outros aspectos, porém é importante ressaltar aqueles que são fundamentais para observar que, diante de tamanha diversidade sociocultural, é um mito afirmar que o PB é uma língua única e não “várias” (MATTOS E SILVA, 2004b.)

### **3.2.1 Os povos e as línguas que aqui habitavam e que aqui chegaram**

Comumente, ao falar em diversidade linguística no Brasil, nos deparamos com noções equivocadas baseadas em construções ideológicas que fazem distinção entre língua nacional, língua de civilização e línguas que não possuem tradição nos estudos acadêmicos (FRANCHETO, 2020). A ideia de uma só nação, uma só língua em oposição ao contexto multilíngue com línguas de tradição mais oral que escrita gerou um forte estigma sociocultural sob as línguas que não são de origem indo-europeia, como as línguas indígenas e as línguas africanas.

Com o objetivo de ampliar a desmistificação de que as línguas indígenas e as línguas africanas são línguas simplificadas em relação ao português no que tange à estrutura, apresentamos nesta sessão um breve histórico a respeito dos povos indígenas brasileiros e algumas de suas línguas, bem como os povos e as línguas africanos que aqui chegaram, ambos diretamente ligados à formação do PB, porém discriminados e diminuídos por ideologias dominantes.

### 3.2.1.1 As nações indígenas e suas línguas

Segundo dados do Censo do Instituto Brasileiro de Geografia e Estatística (IBGE, 2010) e da Fundação Nacional do Índio (FUNAI), nos dias atuais, existem cerca de 274 a 280 línguas indígenas em todo o território brasileiro. No entanto, estudos linguísticos e sócio-históricos (RODRIGUES, 1993; MELATTI, 2014; D'ANGELIS, 2019; dentre outros) apontam que o número de línguas indígenas no Brasil gira em torno de 150 a 180 línguas. Essa diferença numérica ocorre devido ao tipo de metodologia das pesquisas. O questionário aplicado pelo Censo é caracterizado pelo método de autodeclaração, no qual os informantes respondem quais línguas falam e muitos se dizem falantes de certas línguas indígenas por questões de reafirmação de identidade quanto à ancestralidade e pertencimento territorial, mas de fato não falam tais línguas. Nessa disparidade numérica sobre quantas línguas são faladas no Brasil atualmente, podemos questionar quantas eram, então, as línguas indígenas brasileiras no período colonial?

Rodrigues (1993) fez duas projeções sobre a quantidade de línguas existentes no Brasil colonial, utilizando descrições e relatos registrados por Cardim datados de 1584 (1925) e Menéndez com registros do século XVIII (1982) sobre esse período do país. Cardim (1925) fez um levantamento das nações indígenas do Brasil e quantas línguas eram faladas no litoral brasileiro.

Fazendo uma projeção direta do número de línguas achadas na área menor para a área maior, isto é, admitindo que 69 representasse 6.4% do número de línguas existentes em todo território brasileiro, obtemos o número de 1.078 línguas (RODRIGUES, 1993, p.90).

Em Menéndez (1982) há registros antropológicos que documentam cerca de 61 povos indígenas em uma área do interior da Amazônia entre os rios Tapajós e Madeira, do início do século XVIII, quando esse território era praticamente isolado. Utilizando esses dados, Rodrigues (1993) fez uma outra projeção para todo o território brasileiro da existência de 1.273 línguas indígenas no período colonial. O autor faz uma média aritmética com os resultados das duas projeções e propõe o cálculo final de 1.175 línguas. D'Angelis (2019) faz uso da mesma metodologia de Rodrigues (1993) para aplica a novos documentos do Brasil colonial e faz uma nova projeção contando 1.295 línguas indígenas.

Ao fazer uma média aritmética com os resultados dos estudos anteriores, D'Angelis aponta hipoteticamente a existência 1.215 línguas. O genocídio de vários povos indígenas foi o fato histórico determinante para a extinção de mais de 1.000 línguas no Brasil. Dentre várias consequências desastrosas, a perda da história e da cultura dos povos é imensurável.

A redução de 1200 para 180 línguas indígenas nos últimos 500 anos foi o efeito de um processo colonizador extremamente violento e continuado, o qual ainda perdura, não tendo sido interrompido nem com a independência política do país no início do século XIX, nem com a instauração do regime republicano no final desse mesmo século, nem ainda com a promulgação da “Constituição Cidadã” de 1988. (RODRIGUES, 2005, p. 36)

Apesar dessa grande diversidade linguística, os colonizadores portugueses se interessaram em aprender, estrategicamente, o Tupi da Costa que era a língua falada por cerca de dez nações, segundo Cardim (1584 [1925]), ou seja, era a língua mais falada no litoral brasileiro. Conforme os registros coloniais feitos por Cardim, não eram somente os povos do tronco Tupi que falavam essa língua, fato que evidencia mais uma vez a diversidade linguística indígena e que desmitifica a visão simplista de que as nações indígenas só falavam as línguas provenientes do seu tronco linguístico.

Melatti (2014), afirma que o litoral brasileiro era uma área de conflito por conquistas territoriais, e aponta que os tupis eram povos genuinamente guerreiros, conquistavam muitas outras nações indígenas e impunham não somente o domínio territorial, mas também o domínio linguístico. Mesmo havendo variação linguística entre as regiões, o Tupi da Costa era compreendido entre várias nações do litoral do Brasil, era uma língua de contato entre os indígenas, pois muitos povos tinham sua língua materna e falavam o Tupi da Costa como L2. O Tupi da Costa foi a primeira língua aprendida e registrada pelos jesuítas, inclusive foi a língua que deu origem às línguas gerais a partir do contato com o PE.

Os jesuítas descreveram o Tupi da Costa<sup>14</sup> baseando-se nas categorias gramaticais do Latim, dividindo as palavras em classes, atribuindo-lhes significados e funções. Essa descrição trouxe um aspecto marcante no que tange ao entendimento das línguas indígenas: a comparação estigmatizada com o português. Descrever e comparar língua de tipologias diferentes, supervalorizando uma em detrimento da

---

<sup>14</sup> Algumas gramáticas escritas pelos jesuítas foram: A arte de gramática da língua mais falada na costa do Brasil (1595) – Pe. José de Anchieta; Arte da língua brasílica (1621) – Pe. Luis Figueira; Arte da grammatica da língua Brasília da nação kiriri (1699) - Luís Vincêncio Mamiani (RODRIGUES, 2005).

outra, gera não somente preconceito linguístico, mas também uma acentuada desvalorização sociopolítica.

Apesar de ser comum o termo *língua geral*, no singular, Rodrigues (1996, 2000) afirma que foram duas as línguas gerais no Brasil colônia: a língua geral paulista e a língua geral amazônica. A língua geral paulista surgiu na região Sul do Brasil colônia, na capitania de São Vicente, em meados de 1532, chegando a Minas Gerais e Mato Grosso com as expedições dos bandeirantes. A língua geral amazônica originou-se no século XVII, na região norte do país, a Região Amazônica – estendendo-se pelo Maranhão, Pará, Vale do rio Amazonas, parte da Colômbia e Venezuela.

O uso de uma língua geral foi estimulado pelos portugueses com o intuito de estabelecer uma política de homogeneização e consolidar a dominação da população autóctone. O que Borges (2003) chamou de *tupinização da diversidade étnica da colônia* encontrou espaço fértil justamente no “complexo de relações étnicas e linguísticas (miscigenação, inclusive) entre índios e soldados e entre homens e mulheres, causando um impacto sociolinguístico” (BORGES, 2003, p. 117).

As línguas gerais se caracterizaram como a principal forma de comunicação entre índios, negros e afrodescendentes e brancos por cerca de três séculos. Porém, o extermínio da população dos povos indígenas com guerras e constantes epidemias, o intenso tráfico de africanos escravizados, o caótico processo de civilização do Brasil com a institucionalização do português como língua oficial da colônia e a proibição do uso das línguas gerais (decreto oficializado pelo Marquês de Pombal, em 1757), fizeram com que as línguas gerais entrassem em desuso.

A administração central da do Brasil colônia era estabelecida na Bahia, região central da costa. Era natural que a concentração de portugueses fosse maior na região administrativa do que nos extremos Sul e Norte da faixa costeira. Em virtude disso, o contexto de mestiçagem da Bahia era distinto das outras regiões, tendo em vista que o domínio português impossibilitava o uso de línguas indígenas ou de línguas gerais (RODRIGUES, 1996).

Além do domínio português, ocorreram, na Bahia, guerras que extinguiram inúmeros povos indígenas. No território em que, hoje, situam-se os municípios de Poções e Santo Antônio de Jesus, localidades abalisadas neste trabalho, não foi diferente. Ali, os aldeamentos constituíam uma extensa área territorial, conhecida



como Sertão da Ressaca<sup>15</sup>, que vai das margens altas do Rio do Pardo até o médio Rio das Contas (MENDES e ALMEIDA, 2008).

As nações indígenas presentes nessa região eram compostas por índios do tronco Tupi, predominantemente os tupinambás, e do tronco Macro-Jê, como os mongoiós, os aimorés e os pataxós. A ocupação portuguesa foi realizada com a derrota dessas nações indígenas em confrontos violentos que perduraram de meados de 1730 a 1803<sup>16</sup>. Rodrigues (1996) supõe que, diante desse cenário, não houve condições sociolinguísticas para o surgimento de uma língua geral, como ocorreu em São Vicente e na região amazônica.

Confrontando essa suposição de Rodrigues, Argolo (2013, p.79) diz que

o pesquisador Permínio Ferreira encontrou, no Arquivo Público do Estado da Bahia (APEB), um documento de 1794, referente à Vila de Olivença, na então Capitania de Ilhéus, situada ao sul da Bahia, no qual Antônio da Costa Camelo é requerido no sentido de prover Manuel do Carmo de Jesus no cargo de Diretor de Índios, alegando como principal razão para tal o fato "(...) de ser criado naquela vila e saber a língua geral de índios para melhor saber ensinar" (Lobo et al., 2006: 609, grifo nosso).

Argolo (2013) comprova com documentos históricos a existência de uma língua geral falada no Sul da Bahia e propõe uma revisão dos fatos passados na sócio-história do Brasil. O autor diz que dados demográficos da Vila de Olivença, em Ilhéus, apresentados por Lobo (2006), "não existiu, ali, uma população mameluca relevante, mas, ainda assim, houve a formação da língua geral" (ARGOLO, 2013, p.81).

Notadamente, essa assertiva de Argolo (2013) contribui para o aprimoramento dos estudos linguísticos no território baiano. A presença do elemento indígena fora tão drasticamente reduzida, sobretudo, no Sertão da Ressaca<sup>17</sup>, que não nos permite fazer inferências sobre até que ponto a língua geral teria perpassado pela faixa costeira e influenciado a formação do português do interior da Bahia em aspectos

---

<sup>15</sup> O nome Sertão da Ressaca pode ter duas origens: (i) fenômeno de invasão das águas sobre o sertão ou (ii) derivado da palavra *ressaco* que significa *funda baía de mato baixo circundada por serras* (MENDES e ALMEIDA, 2008).

<sup>16</sup> Por volta de 1730, o bandeirante João da Silva Guimarães foi o encarregado da administração portuguesa para desbravar a região do Sertão da Ressaca. João da Silva Guimarães juntamente com João Gonçalves da Costa, criou aliança com os mongoiós para derrotar os aimorés, que eram povos rivais. Após a derrota dos aimorés, Gonçalves da Costa liderou o massacre de seus aliados em uma emboscada que ficou conhecida como o "banquete da morte". Os povos pataxós, que eram em menor número, fugiram para se refugiar em regiões mais extremas ao que, hoje, é o Sul do Estado da Bahia (SILVA e SOUSA, 2013).

<sup>17</sup> Vale salientar que a Vila de Olivença é localizada mais ao extremo Sul do estado, enquanto o Sertão da Ressaca compreende mais o Sudoeste.

gramaticais e estruturais, para além do léxico, na mesma proporção que as línguas africanas.

Segundo Mattos e Silva (2004a), o processo “etnocida” sofrido pelos povos indígenas, os índios integrados brasileiros não tiveram possibilidade de serem os principais difusores do português no Brasil. Somado a isso, a presença de africanos e seus descendentes se tornou cada vez mais intensa e duradoura no interior do estado, ora escravizados, ora fugitivos que formaram quilombos na tentativa de articular algum modo de sobrevivência (MATTOS E SILVA, 2004a).

O Quadro 3, com dados apresentados por Mussa (1991), mostra a extrema redução dos povos indígenas, que perfazia 50% da população do Brasil colônia, no século XVI, e, ao final do século XIX, não passava de 2%.

Quadro 3. Distribuição demográfica da população no Brasil colonial.

<b>Etnias</b>	<b>1538-1600</b>	<b>1601-1700</b>	<b>1701-1800</b>	<b>1801-1850</b>	<b>1851-1890</b>
Africanos	20%	30%	20%	12%	2%
Negros brasileiros	-	20%	21%	19%	13%
Mulatos	-	10%	19%	34%	42%
Branços brasileiros	-	5%	10%	17%	24%
Europeus	30%	25%	22%	14%	17%
Índios integrados	50%	10%	8%	4%	2%

Fonte: Adaptado de Mussa (1991, p. 163)

Como já mencionado, o extermínio da população indígena no Brasil, além de outras inúmeras perdas culturais, levou à morte milhares de línguas. Dessa forma, esta pesquisa centra-se nos possíveis efeitos de contato da língua portuguesa com as línguas africanas na formação do português popular do interior da Bahia. As informações do Quadro 3 sobre a proporção das etnias no Brasil remetem a um aspecto importante levantado por Lucchesi (2019). Segundo ele, um dos motivos pelos quais não houve uma crioulização do português no Brasil seria a diminuição expressiva da proporção de falantes da língua de superestrato, fato que teria tornado restrito o acesso dos falantes do substrato aos modelos da LA. Lucchesi (2019) ainda acrescenta que

além da proporção absoluta entre o grupo dominante e o grupo dominado, a estrutura da sociedade colonial brasileira também parece diferir da estrutura prototípica das sociedades de plantação do Caribe. Como toda sociedade de plantação, a sociedade açucareira do Nordeste do Brasil estruturou-se em torno do engenho, a grande propriedade latifundiária “que, além de concentrar a produção e vinculá-la ao seu sentido mercantil, tornava possível executar as funções que dão sentido a uma série de atividades” (FERLINI, 2002, p. 26) (LUCCHESI, 2019, p. 234).

Na seção a seguir, apresento os principais fatos sócio-históricos e sociolinguísticos acerca dos povos africanos que foram trazidos para o Brasil.

### 3.2.1.2 Os povos africanos e suas línguas

Estima-se que cerca de quatro milhões de africanos foram escravizados e transportados durante três séculos de tráfico (Schwarz & Starling, 2015, p.82). Considerando Mufwene (2018), as línguas se movem junto com os povos que as falam, logo junto com os povos africanos escravizados aportaram, no Brasil, suas línguas. Segundo Petter (2006), aproximadamente duzentas a trezentas línguas africanas foram transplantadas para o Brasil durante o tráfico negreiro. A intensa presença de africanos no Brasil se caracterizou como um dos fatos históricos mais marcantes na formação da sociedade brasileira.

Algumas questões importantes devem ser destacadas: Como a língua portuguesa se tornou hegemônica em meio à diversidade de línguas indígenas e africanas, principalmente, com as quais teve contato? Por que nenhuma das trezentas línguas africanas permaneceu no quadro linguístico do Brasil?

As causas do desaparecimento da maioria das línguas indígenas estão diretamente ligadas ao genocídio dos povos indígenas. Quanto às línguas africanas, as causas são incertas. Faraco (2016, p. 145) aponta que, possivelmente, em contextos de contato massivo de línguas diversas, os falantes tendem a abandoná-las na tentativa de criar condições de construir interações. A diversidade, nesse caso, levaria à uniformização. No entanto, a falta de dados torna essas causas apenas especulativas.

Mattoso (1982, p. 22-23) resume os principais períodos do tráfico negreiro em quatro grandes ciclos:

- a) O ciclo da Guiné, no século XVI, no qual houve a predominância da escravização de sudaneses;
- b) O ciclo do Congo e de Angola, no século XVII, marcado pela escravização de africanos bantos;
- c) O ciclo da Costa de Mina, no século XVIII, caracterizado pela presença de sudaneses escravizados, novamente;
- d) E, impulsionado pelo cultivo do café, no século XIX, houve a importação de africanos de todas as regiões, mas principalmente originários de Angola ou de Moçambique.

Angola e Moçambique se configuraram como pontos estratégicos de ocupação por se localizarem na costa do continente africano. Isso favoreceu o deslocamento das tropas marítimas portuguesas pelo Atlântico, principalmente para o Brasil.

Petter (2006, p. 124) menciona que, de acordo com a região de origem dos africanos trazidos para o Brasil, as línguas africanas que aqui chegaram são originárias de duas grandes áreas, essencialmente:

- a) área oeste africana: com um grande número de línguas de grupos tipologicamente diversos: (i) “oeste-atlântica” (*fulfulde, wolof, serer, temne etc.*); (ii) “mandê”; (iii) “kwa” ou “gbe” (*even, gen, aja, fon*); (iv) “benuê-congo”; (v) “chádicas” (*haussá*); e (vi) “nilo-saariana”;
- b) área banta: com um número menor de línguas, mas faladas por um número maior de escravizados: (i) *quicongo*, falada pelos bacongos, numa região que corresponde ao antigo Reino do Congo; (ii) *quimbundo*, falada pelos ambundos, na região central de Angola, antigo reino de Ndongo; (iii) *umbundo*, falada pelos ovibundos, na região de Benguela, em Angola.

É importante destacar que o contato entre línguas africanas de grupos diferentes e também com o português ocorrera antes destas aportarem no Brasil, visto que o domínio português em território africano aconteceu antes da chegada dos portugueses em terras brasileiras, como descrito brevemente na seção 3.2. Muitos africanos trazidos para o Brasil possuíam algum conhecimento básico de português, principalmente aqueles vindos da atividade agrícola estável nas ilhas de Cabo Verde e de São Tomé, sendo um tipo de elite trabalhadora técnica (cf. FARACO, 2016, p. 144).

Stuart Schwartz (1988) estudou como os primeiros africanos se estabeleceram nos engenhos do Recôncavo Baiano, uma das regiões pioneiras no estabelecimento de uma longa economia açucareira. Segundo aquele autor, eles foram trazidos provavelmente das ilhas atlânticas portuguesas, onde já se cultivava o açúcar em regime de plantation, para ocupar posições especializadas no processo produtivo. Enquanto a mão de obra indígena ainda constituía a força de trabalho dos canaviais, esses primeiros africanos vieram para operar o engenho propriamente dito, a casa da moenda, a área de transformação da matéria-prima, como mestres de açúcar, tanoeiros, ferreiros e outras ocupações que exigiam experiência e treinamento. Podem ser considerados uma espécie de elite da classe trabalhadora atlântica, no sentido de que tocavam um dos ramos mais avançados da indústria ocidental do século XVI (REIS, 2007, p. 82).

Além disso, a política de Portugal se caracterizava por misturar diferentes etnias africanas para impedir a formação de motins. Os africanos eram misturados ao serem capturados, no processo de espera para o embarque nos navios e na viagem pelo Atlântico (PETTER, 2006, p. 124-125).

Em Angola, os portugueses utilizaram uma variante desse sistema, um navio ancorado permanentemente num porto, que oferecia a vantagem de conter com maior segurança a concentração de um grande número de cativos. Os escravos, submetidos a essa organização particular de tráfico, participaram de uma situação linguística nova, transitória, mas que pela sua duração e renovação regular, pôde influenciar a situação linguística no Brasil (BONVINI; PETTER, 1998, p. 74). Essa situação se caracterizava pela concentração forçada e prolongada de locutores de línguas africanas diferentes, embora tipologicamente próximas, que pôde conduzir, no caso de Angola, à adoção do quimbundo (falado em Luanda e ao longo do rio Kwanza até sua foz) como língua veicular. Por outro lado, nesse mesmo período, ocorreu um contato forçado e prolongado com a língua portuguesa; primeiramente com aquela utilizada pelos “pombeiros”, recrutadores de cativos no interior das terras; a seguir, ao longo da costa africana, com os “negreiros”, brasileiros, que seriam talvez seus futuros senhores. Para muitos escravos originários de Angola, o multilinguismo encontrado no Brasil, resultante do convívio de línguas africanas e português, não será inédito, o que poderá explicar tanto o emprego de uma língua veicular africana quanto o uso do português falado pelos senhores (PETTER, 2006, p. 126).

De forma cruel, os africanos foram separados de suas famílias, retirados de suas terras nativas, forçados a conviverem em grupos distintos em condições subumanas e escravizados pelos colonizadores portugueses para servirem aos senhores da sociedade brasileira. Nesse contexto, um grupo de africanos falava várias línguas e, como precisavam se comunicar, assimilavam de forma rudimentar a língua dominante, o português.

Segundo Mussa (1991), os africanos utilizavam o português como língua de intercurso com a sociedade branca, mas quando existiam grupos colíngues era mantida a língua materna de seus membros. Apesar disso, o autor afirma que é

duvidoso postular sobre a existência de uma língua geral africana, mesmo na zona do garimpo, em que se tem registros da utilização de um dialeto da língua *fon* no século XVIII.

Pessoa de Castro (2006) destaca que dos quatro milhões de africanos trazidos para o trabalho escravo no Brasil, cerca de 75% eram falantes de línguas do grupo banto. Segundo a autora, a região banto compreende um grupo de 300 línguas muito semelhantes, faladas em 21 países. Dentre essas línguas, as que tiveram maior número de falantes no Brasil foram o quicongo, o quimbundo e o umbundo, grafados originalmente como *kikongo*, *kimbundu* e *umbundu*, respectivamente.

Tamanha era a presença de falantes dessas línguas no Brasil, sobretudo na Bahia, que o padre Pedro Dias escreveu “A arte da língua de Angola”, uma gramática publicada em 1687, em Lisboa, como meio de instruir os jesuítas e facilitar o trabalho de catequese dos “25 mil etíopes” africanos, que se encontravam na cidade do Salvador sem falar português (cf. PESSOA DE CASTRO, 2006, p. 6).

No território baiano, também se destacaram línguas oriundas da região oeste-africana, as do grupo *kwa*, em especial o iorubá<sup>18</sup> e as do grupo *ewe-fon*. No entanto, “as línguas banto foram as mais importantes no processo de configuração do perfil do português brasileiro, devido à antiguidade e superioridade numérica de seus falantes e a grandeza da dimensão, no tempo e no espaço, alcançada pela sua distribuição humana no Brasil colonial” (PESSOA DE CASTRO, 2006, p 118).

Reis (2007, p. 81) diz que, inicialmente, os africanos escravizados foram trazidos para atuar na economia açucareira, mas a força de trabalho africana se estendeu em praticamente todos os setores da sociedade: na agricultura de exportação, na agricultura de abastecimento interno, na criação de gado, no trabalho doméstico, nos ofícios mecânicos e em toda ordem de ocupações urbanas, nas ruas de Recife, Salvador, Ouro Preto, Rio de Janeiro, São Paulo, Porto Alegre e outros centros urbanos, principalmente com o avanço do processo de civilização do Brasil marcado pela vinda da Família Real, em 1808.

Usando a expressão “multidão sem voz” da historiadora Kátia Mattoso para se referir aos africanos e aos afrodescendentes brasileiros, Mattos e Silva (2004a)

---

<sup>18</sup> Segundo Pessoa de Castro (2006), devido a uma introdução tardia e à numerosa concentração dos seus falantes na cidade do Salvador, é possível encontrar aportes do iorubá nos dias atuais, ainda que seu uso seja mais identificado no âmbito religioso do candomblé.

conclui que foram eles os principais difusores do português popular pelo território brasileiro, sobretudo na Bahia e em Pernambuco.

Contrariamente ao que ocorreu com os indígenas [cujo número decresceu drasticamente], os africanos e os afrodescendentes estão no patamar de 60% da população do Brasil entre os séculos XVII ao XIX. Tendo de abdicar de suas línguas de origem, como referido, não tinham eles escolha: tiveram de aprender, num processo de transmissão linguística irregular – na designação da crioula atual – a língua da colonização. Certamente, junto com o pequeno contingente de indígenas integrados ao processo colonizador, são eles que vão dar forma ao português geral brasileiro, antecedente, como penso, do português popular ou vernáculo (MATTOS E SILVA, 2004a, p. 128).

No que diz respeito ao estudo do português afro-brasileiro no Estado da Bahia, os africanos que chegavam à colônia de Leopoldina (Helvécia) eram de origem distinta, como registram Baxter e Lucchesi (1999, p. 131, *apud* Lucchesi 2002, p. 82), mas eram em sua maioria falantes de línguas dos grupos banto e, em algumas fazendas, do grupo Kwa. De acordo com os autores, a quantidade de suíços, alemães, principalmente franceses e brasileiros na Colônia não era significativa, representava 10% da população. Os 90% restantes eram compostos por: 50% de africanos e 40% de afro-brasileiros.

Foi investigando o português falado em comunidades afrodescendentes rurais semi-isoladas do interior da Bahia (Helvécia, Sapé, Cinzento e Rio de Contas) que Baxter (1992), Baxter e Lucchesi (1997) e Lucchesi e Baxter (2009) concluíram que a influência das línguas africanas não se fixou somente no léxico. Segundo esses autores, o comportamento morfossintático do português falado nessas comunidades indica influência significativa das línguas africanas no processo de formação do que eles chamaram português afro-brasileiro. Características como: (i) redução das marcas morfológicas flexionais, ocasionando uma variação na concordância nominal e verbal; (ii) dupla negação; (iii) monomorfismo pronominal; (iv) construções ativas de sentido passivo etc. têm sido apontadas como decorrentes efeitos do processo do contato linguístico.

Considerando as condições sócio-históricas do povoamento do Brasil, e da Bahia, em que os povos colonizados e escravizados se concentraram em regiões interioranas ou nas periferias das cidades, apresento brevemente na subseção a seguir como o processo de urbanização do Brasil levou a uma polarização linguística e à existência da dicotomia: português rural *versus* português urbano.

### 3.2.2 Os efeitos do contato na urbanização do Brasil e a dicotomia *rural-urbano*

Com a diminuição do tráfico de africanos para trabalhar como mão de obra escrava, por volta de 1850, e com a abolição da escravatura, em 1888, o processo de imigração de europeus passou a ser estimulado pelo governo brasileiro como estratégia econômica para ocupar mais regiões do país, além de estimular o “branqueamento” da população, já que a miscigenação era vista pelos portugueses como uma barreira ao desenvolvimento econômico e social do Brasil (KREUTZ, 2000).

Lucchesi (2002) relata que a necessidade de mão-de-obra nas lavouras de café, basicamente composta por africanos e afrodescendentes, intensificou o processo de imigração, no qual aproximadamente mais de três milhões de imigrantes chegaram ao Brasil. Ainda segundo Lucchesi (2002), os imigrantes chegados ao Brasil inicialmente ocuparam os estratos mais baixos da hierarquia social, porém, em virtude de racismo estrutural, ascenderam rapidamente na escala social brasileira, enquanto os africanos e afrodescendentes permaneceram nas classes mais inferiores.

Com a finalidade de fixar a hegemonia da língua portuguesa no Brasil, o governo criou inúmeros decretos, que garantiam o ensino da língua portuguesa no país. Logo, as línguas dos imigrantes tiveram uso restrito nas pequenas às suas pequenas colônias ou grupos de falantes, e permanece quase que de igual modo em algumas regiões do Brasil, principalmente no sul do país (KREUTZ, 2000). Mattos e Silva (2004a) resume o contexto multilíngue que formou o PB da seguinte forma:

[...] O português brasileiro descende do europeu e, no Brasil, tomou a sua forma na complexa interação entre a língua do colonizador e, portanto, do poder e do prestígio; as numerosas línguas indígenas brasileiras; as também numerosas línguas africanas chegadas pelo tráfico negreiro[...]; as línguas que emigraram para o Brasil da Europa e da Ásia [...]. Dessa potencial Babel linguística, foi se definindo [...] o formato brasileiro contemporâneo da língua portuguesa (MATTOS E SILVA, 2004a, p. 140)

Considerando essa ‘Babel linguística’, Lucchesi (2001, 2002, 2015) propôs o que chamou de *polarização sociolinguística do Brasil*, uma perspectiva de que a formação do PB, notadamente a variedade popular, está relacionada aos contatos linguísticos ocorridos entre o português e as línguas indígenas e, principalmente, as africanas nas regiões rurais do país. Lucchesi (2001, 2002) considera a oposição entre



as zonas rurais e os centros urbanos como ponto fundamental para a constituição das normas objetivas brasileiras em função da diferenciação étnica, demográfica, sociocultural e econômica dos espaços rurais e urbanos.

De acordo com Lucchesi (2002), de um lado estaria a maioria da população brasileira centrada, até o início do século XIX, nas regiões rurais, desenvolvendo as atividades econômicas que sustentavam a economia colonial. Uma população rural formada, maioritariamente, por negros e mestiços, que adquiriram a língua portuguesa como L2 por meio da oralidade, sem a ação de um mecanismo normatizador e sem acesso à escolarização. Do outro lado, nos espaços urbanos, estaria a minoria da população brasileira, a elite colonial, que buscava preservar os padrões culturais da metrópole, tentando manter a língua portuguesa mais próxima possível do modelo europeu, o que contribuiu para a origem da norma culta do PB em oposição à norma popular, falada por mestiços descendentes de índios e africanos (LUCCHESI, 2002).

Essa oposição sociolinguística rural *versus* urbano também foi estudada por outros autores, como Bortoni-Ricardo (2005), que afirma que a formação do PB popular “[...] foi, possivelmente, mais acentuada a influência do adstrato indígena e do *pidgin* falado pelos negros entre si e nos seus contatos com a população branca e mestiça” (BORTONI-RICARDO, 2005, p. 32).

A rápida e caótica urbanização do Brasil<sup>19</sup> a partir de meados de 1930, que até o século XIX era um país majoritariamente rural, contribuiu ainda mais para a estratificação da realidade sociolinguística brasileira, compreendida por Bortoni-Ricardo (2005; 2011) como um *continuum* de urbanização. Constituído por dois polos extremos, as variedades rurais mais isoladas e as variedades urbanas cultas, esse *continuum* seria formado também por zonas intermediárias ocupadas por traços contínuos e gradativos mais ou menos próximos de um dos extremos, sendo possível situar um falante em qualquer posição do *continuum* (BORTONI-RICARDO, 2005, p. 40).

Lucchesi (2015) aponta que a ideia de *continuum* proposta por Bortoni-Ricardo (2005) complementa a perspectiva da polarização sociolinguística e acrescenta uma

---

<sup>19</sup> Os processos de urbanização e industrialização no Brasil, no século XX, foram impulsionados pela Revolução Industrial. Segundo dados de censos demográficos do IBGE, a taxa de urbanização do país foi de 31, 1% em 1940 para 84, 4% em 2010. Um processo acelerado da urbanização, típico dos países de industrialização tardia, o que provocou profundas alterações no cenário socioeconômico brasileiro, tais como a expansão da malha rodoviária e dos meios de comunicação de massa, a mecanização do campo, a democratização do ensino básico e intensos fluxos migratórios da população do campo em direção aos centros urbanos.

correlação da oralidade-letramento e da monitoração estilística ao modelo da polarização.

[...] aplicando o modelo dos três continua ao quadro de polarização, têm-se, no ponto extremo da norma popular, as comunidades rurais mais isoladas do interior do país, cujos falantes, imersos na oralidade, exibiriam um espectro de variação estilística bastante reduzido, já que é escasso também o seu contato com as demais variedades do português. No outro extremo da norma culta, estariam os falantes dos grandes centros urbanos, com alto grau de escolaridade, cujo extenso repertório linguístico amplia bastante o espectro de sua variação estilística. Entre um extremo e outro, vislumbra-se um continuum de níveis intermediários. À medida que se passa do pólo extremo da norma popular para a norma culta, cresce progressivamente o grau de urbanização e letramento, bem como o espectro da variação estilística. Assim, vão se descortinando sucessivos cenários sociolinguísticos que podem ser assim discriminados, considerando a escala *rural > rurano > urbano* (LUCCHESI, 2015, p. 150).

Considerando essa noção de *continuum*, esta pesquisa também objetiva verificar em qual cenário sociolinguístico da escala *rural > rurano > urbano* encontra-se o português do interior da Bahia discutindo o *status* do fenômeno variável da CN.

Associados os principais aspectos sócio-históricos e sociolinguísticos à formação do PB, infiro que a variedade do português popular do interior da Bahia tenha sido formada por uma população de idioletos com um conjunto de variantes do PE e das línguas africanas, principalmente, conforme pressupõe a hipótese da evolução linguística por meio da competição e seleção de traços (MUFWENE, 2001; 2008). Após a análise dos dados (cf. Capítulo 5), desenvolvo detalhadamente, no Capítulo 6, essa inferência e apresento a proposta de tese deste trabalho.

### 3.3 SÍNTESE

Mediante os objetivos descritos na seção 3, apresentei as principais linhas teóricas que tratam o contato entre línguas como fator que impulsiona desde mudanças linguísticas até a formação de novas línguas. Seguindo os pressupostos da variação e da mudança linguísticas associadas a fatos sócio-históricos, fiz uma sistematização dos principais aspectos da história do povoamento do Brasil – o genocídio dos povos indígenas e a escravização de africanos -, sobretudo da Bahia, relacionando os povos que aqui habitavam e os que aqui chegaram com suas línguas.

Diante disso, discorri sobre a formação da sociedade brasileira quanto à urbanização, considerando a proposição de Bortoni-Ricardo (2005) sobre um *continuum* do PB, em que nas extremidades há dois lados, um com as variedades rurais mais isoladas, e o outro com as variedades urbanas cultas mais presentes nos centros urbanos. Para aprofundar as discussões sobre isso, mencionei a proposta reelaborada de Lucchesi (2015) que reconhece o modelo do *continuum* como complementar ao da polarização do PB, tendo em vista a origem rural do *corpus* deste trabalho.

Oportunamente nos Capítulos 5 e 6, teço considerações sobre os resultados desta pesquisa em relação ao contato com as línguas bantu. Antes disso, porém, é necessário apresentar os procedimentos metodológicos envolvidos neste trabalho, aos quais se referem o Capítulo 4, a seguir.

## 4. ASPECTOS TEÓRICO-METODOLÓGICOS

Como já mencionado, este trabalho investiga a variação no uso da CN no âmbito do SN do português falado no interior do Estado da Bahia com o objetivo de descrever essa variedade linguística, observando a ação de fatores estruturais e sociais, levando em consideração a possibilidade de o contato linguístico ter contribuído para a ocorrência de determinadas estruturas características do PB. Para o desenvolvimento do trabalho, adoto os pressupostos da Sociolinguística Variacionista (LABOV, 2008[1972]). Considerando o objeto de estudo desta pesquisa, a variação no uso da CN, a utilização dessa teoria se justifica porque estabelece que a variação é inerente à língua, salientando que é uma variação ordenada, não aleatória, condicionada por fatores linguísticos e sociais.

Neste Capítulo, faço algumas considerações sobre a Sociolinguística Variacionista e seus procedimentos de análise de dados, discorro sobre a origem e o tratamento dos dados da CN utilizados neste trabalho, destacando as comunidades de Poções e de Santo Antônio de Jesus, das quais os dados foram extraídos. Faço o detalhamento das variáveis (dependente e independentes) selecionadas para a análise dos dados do trabalho. Por fim, exponho os aspectos metodológicos, bem como o programa computacional R, utilizado para a análise estatística dos dados.

### 4.1 A SOCIOLINGUÍSTICA

A Sociolinguística surgiu no final da década de 1960 com a proposta de analisar os aspectos sociais e culturais que motivam as práticas linguísticas, buscando preencher lacunas deixadas pelos estudos formalistas (Labov, 2008 [1972]), como o Estruturalismo, acerca da questão da mudança linguística. O cerne das pesquisas sociolinguísticas é o estudo da fala, dando um novo rumo às investigações linguísticas que, até então, focalizavam somente na estrutura linguística.

A base do conhecimento intersubjetivo na lingüística tem de ser encontrada na fala – a língua tal como usada na vida diária por membros da ordem social, este veículo de comunicação com que as pessoas discutem com seus cônjuges, brincam com seus amigos e ludibriam seus inimigos. (WEINREICH, LABOV E HERZOG (2006 [1968], p. 13).

Ao discorrer brevemente sobre a sociolinguística e o seu objeto de estudo – os padrões de comportamento linguístico observáveis dentro de uma comunidade de fala (Labov, 2008 [1972]) – não anulo o fato de que as teorias estruturalistas reconheciam, de certa forma, o caráter variável da língua, no entanto, por questões metodológicas, a preferência era restringir os trabalhos à análise da *langue* em detrimento da *parole*, tal como o fez Ferdinand de Saussure (1995 [1916]).

Para ele, a *langue* “um produto social da faculdade da linguagem e um conjunto de convenções necessárias, adotadas pelo corpo social para permitir o exercício dessa faculdade do indivíduo”, enquanto a fala é conceituada como “as combinações pelas quais o falante realiza o código da língua no propósito de exprimir seu pensamento pessoal; o mecanismo psicofísico que lhe permite exteriorizar essas combinações” (SAUSSURE, 1995 [1916], p. 17; 22). De acordo com o linguísta, a *langue* e a *parole* estão associadas e interligadas.

a língua é necessária para que a fala seja inteligível e produza todos os seus efeitos; mas esta é necessária para que a língua se estabeleça; historicamente, o fato da fala vem sempre antes. [...] Existe, pois, interdependência da língua e da fala; aquela é ao mesmo tempo o instrumento e o produto desta. Tudo isso, porém, não impede que sejam duas coisas absolutamente distintas. (SAUSSURE, 1995 [1916], p. 27)

Além de assumir que a língua e a fala estão relacionadas, Saussure diz que a língua possui um aspecto social e homogêneo e a fala um aspecto individual não sistematizável, ou seja, na concepção saussuriana não se poderia enquadrar a fala como um sistema ordenado. Dessa forma, Saussure opta por estudar somente a estrutura da língua em si.

Pode-se, a rigor, conservar o nome de Linguística para cada uma dessas duas disciplinas e falar duma Linguística da fala. Será, porém, necessário não confundi-la com a Linguística propriamente dita, aquela cujo único objeto é a língua. Unicamente desta última que cuidaremos, e se por acaso, no decurso de nossas demonstrações, pedirmos luzes ao estudo da fala, esforçar-nos-emos para jamais transpor os limites que separam os dois domínios. (SAUSSURE, 1995 [1916], p. 28)

A Sociolinguística, então, rompe com esse modelo e incorpora aos estudos linguísticos os aspectos sociais e culturais que motivam as práticas linguísticas. Weinreich, Labov e Herzog (2006[1968]) lançam a obra *Empirical foundations for a*

*theory of language change* (Fundamentos empíricos para uma teoria da mudança linguística), na qual postulam as correlações entre os fatores linguísticos e os fatores sociais com a finalidade de compreender que a língua é heterogênea, ou seja, varia e a variação linguística é necessária para o funcionamento real de uma língua. Além disso, os autores afirmam que a língua muda porque varia, e que o processo de variação linguística nem sempre resulta em mudança linguística, porém toda mudança linguística indica que houve um estágio de variação na estrutura linguística.

Certamente não basta apontar a existência ou a importância da variabilidade: é necessário lidar com os fatos de variabilidade com precisão suficiente para nos permitir incorporá-los em nossas análises da estrutura linguística. (WEINREICH, LABOV & HERZOG, 2006[1968], p. 107).

Por levar em consideração o aspecto de variabilidade da língua em situações reais de uso e buscar descrever a língua a partir de seus condicionamentos sociais e linguísticos, a Sociolinguística também é chamada de Teoria da Variação e da Mudança Linguística. Labov (2008 [1972]) apresenta um modelo teórico-metodológico que se dedica a estudar o uso vernacular a língua e as formas em variação dentro de comunidade de fala.

Estudar a variação linguística perpassa por vários vieses, dentre eles, o de estudar as formas em variação, as variantes linguísticas, dentro de uma comunidade de fala. De acordo com Labov (2008 [1972]), variante linguística é cada uma das maneiras de se dizer a mesma coisa em um mesmo contexto, isto é, com o mesmo valor de verdade, e comunidade de fala é um grupo de pessoas que compartilham traços linguísticos que distinguem seu grupo de outros.

O conjunto de variantes é chamado de variável linguística, que pode ser dividida e classificada em variável dependente e variáveis independentes. O fenômeno linguístico a ser estudado é a variável dependente, por exemplo, neste trabalho a variável dependente é aplicação da regra de CN no SN; as variantes são as formas em alternância, por sua vez, são a presença e a ausência da regra de concordância. A utilização de uma determinada variante em detrimento de outra é motivada pelas variáveis independentes, que são os fatores linguísticos ou estruturais e/ou extralinguísticos – sociais.

É através da análise das variáveis independentes que podemos explicar o quadro de variação presente na comunidade de fala. Labov (1983, p. 31) afirma que

as pressões sociais operam continuamente sobre a linguagem, não desde um passado remoto, mas como uma força imanente que atua constantemente no presente. As variáveis sociais, tais como, estilo contextual, *status* e mobilidade social, etnicidade, sexo, idade, nível de escolaridade, entre outras, são levadas em consideração para averiguar o encaixamento social do fenômeno linguístico. O quadro de variação pode ser configurado em duas possibilidades: podemos verificar se ocorre uma *variação estável* ou uma *mudança em progresso*.

Com relação à variação estável, entende-se 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 (LABOV, 2008 [1972]).

O estudo da variação linguística é uma maneira de também estudar a mudança linguística, observando atentamente a influência de questões externas à estrutura linguística. Nessa perspectiva, podemos observar os problemas da mudança linguística, postulados por Weinreich, Labov e Herzog (2006 [1968], p. 121-124): o problema do condicionamento, que diz respeito às condições gerais da estrutura para que a mudança inerente ao sistema ocorra; 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 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.

Como podemos ver, a Sociolinguística é um aparato teórico-metodológico que objetiva dar conta da descrição dos condicionamentos linguísticos e sociais que regem a variação e a mudança linguística. Nas seções seguintes, apresento os métodos que constituem a base empírica desta pesquisa.

## 4.2 A ORIGEM DOS DADOS

Uma das características da Sociolinguística é a pesquisa empírica, ou seja, a compreensão do funcionamento da língua e de seus fenômenos variáveis a partir da análise de dados reais de fala, da observação do vernáculo do falante. Partindo desse pressuposto, decidi estudar a CN na fala de indivíduos do PB popular do interior da Bahia, mais precisamente dos municípios de Poçoões e de Santo Antônio de Jesus.

As amostras de fala que compõem o *corpus* deste trabalho fazem parte do acervo de fala do *Projeto Vertentes do Português Popular do Estado da Bahia*, do qual também fazem parte as amostras de fala do português afro-brasileiro baiano, bem como amostras do português popular de Salvador, capital do estado.

O Projeto Vertentes foi implementado em 2002, na UFBA, encontra-se sob a coordenação do Professor Doutor Dante Lucchesi e do Professor Doutor Gredson dos Santos, e tem como objetivo principal traçar um panorama sociolinguístico do português popular da Bahia, levando em conta alguns aspectos, tais como: a relevância do contato entre línguas na sua formação histórica e os processos de difusão linguística a partir dos grandes centros urbanos.

Para analisar os processos de variação e mudança no âmbito da morfossintaxe do português popular do Estado da Bahia, o Projeto Vertentes segue o enquadramento teórico e metodológico da Sociolinguística Variacionista, com contribuições da Teoria Gerativa da Gramática por alguns de seus membros.

A base empírica fornecida pelo Projeto Vertentes permitiu a realização de diversos trabalhos que contribuem para traçar um panorama sociolinguístico da fala popular do Estado da Bahia. Alguns trabalhos fizeram um mapeamento linguístico de fenômenos variáveis, traçando um *continuum* linguístico desde as comunidades rurais afro-brasileiras, passando pelo interior até à capital do estado (LUCCHESI, 2002, 2009, 2015).

Conforme apresentado no Capítulo 3, segundo Bortoni-Ricardo (2005), o PB deve ser observado “como um *continuum* de urbanização, que se estende desde as variedades rurais geograficamente isoladas [...] até a variedade urbana culta, [...] podendo-se situar um falante em qualquer posição ao longo deste *continuum*.” (p. 40). Mas em estudos anteriores (BORTONI-RICARDO, 1988), a autora cunhou o termo



*rurbano* para caracterizar as populações rurais com um certo nível de integração com a cultura urbana, até mesmo a população urbana que ainda preserva seus antecedentes rurais. Antonino (2007), ao utilizar os *corpora* de Poções e Santo Antônio de Jesus para seu trabalho de dissertação de mestrado, afirmou que

vários falantes 'rurbanos' foram detectados, já que a movimentação entre a sede e a zona rural do município é comum e até intensa. Muitos passaram grande parte da vida na 'roça' e, após determinados eventos, como o casamento ou a aquisição dos proventos da aposentadoria, resolvem ter uma vida 'na cidade', que a maioria das vezes é a sede do seu município. Assim, o intercâmbio cultural e lingüístico é bastante intenso e, para esses falantes, o conceito de 'rurbano' é bastante adequado. (ANTONINO, 2007, p. 39).

Atrelando as perspectivas de Bortoni-Ricardo (1988, 2005, 2011), a observação empírica de Antonino (2007) e a proposta de distribuição da sociedade brasileira proposta por Lucchesi (2015), ilustrada no Quadro 4, esta pesquisa também objetiva verificar em qual cenário sociolinguístico da escala *rural > rurbano > urbano* encontra-se o português do interior da Bahia discutindo o *status* do fenômeno variável da CN.

Quadro 4. Distribuição da sociedade brasileira considerando o *continuum* rural < rurbano < urbano proposta por Lucchesi (2015).

Rural	<ul style="list-style-type: none"> <li>• membros analfabetos ou semianalfabetos de comunidades rurais isoladas</li> <li>• membros analfabetos ou semianalfabetos de comunidades rurais menos isoladas</li> </ul>
Rurbano	<ul style="list-style-type: none"> <li>• habitantes de baixa ou nenhuma escolaridade de pequenas cidades do interior (distinguindo-se imigrados do campo e nascidos na cidade)</li> <li>• habitantes de baixa ou nenhuma escolaridade emigrados do campo para a periferia das grandes cidades</li> </ul>

Urbano	<ul style="list-style-type: none"> <li>• habitantes de baixa ou nenhuma escolaridade da periferia de grandes cidades</li> <li>• habitantes de baixa ou nenhuma escolaridade de bairros populares mais centrais das grandes cidades</li> <li>• trabalhadores e operários qualificados das grandes cidades e centros industriais</li> <li>• comerciantes e trabalhadores técnicos e administrativos das pequenas e médias cidades do interior do país</li> <li>• comerciantes e trabalhadores técnicos e administrativos dos grandes centros urbanos</li> <li>• altos funcionários do judiciário e do executivo, profissionais liberais e intelectuais de pequenas e médias cidades do interior do país</li> <li>• altos funcionários do judiciário e do executivo, profissionais liberais e intelectuais dos grandes centros urbanos</li> </ul>
--------	--

Fonte: LUCCHESI (2015, p. 153).

#### 4.2.1 As comunidades: Poções e Santo Antônio de Jesus

Poções é um município baiano, localizado a 444km da capital do Estado, Salvador, situa-se na região Centro Sul do Estado, na microrregião de Vitória da Conquista. A história do povoamento de Poções data da segunda metade do século XVIII e, segundo Silva (2005), a exploração do território pelo colonizador europeu iniciou-se quando o coronel André da Rocha Pinto tomou rumo pelo Rio de Contas até o Rio Verde e a cabeceira do Rio São Mateus.

Historiadores apontam que na região habitavam os índios mongoiós, ramificação dos Camacãs do grupo Gê. Além da presença indígena, havia a presença de africanos vindos traficados de Angola e Moçambique, assim como de negros nascidos no Brasil. Com a ocupação do território pelos colonizadores, iniciada pelas incursões bandeirantes, foram instaladas fazendas nas quais eram praticadas a pecuária e a agricultura de subsistência, posteriormente foi instaurada a cultura do algodão devido às condições favoráveis da região.

O município alcançou a independência política no ano de 1883, desvinculando-se de Vitória da Conquista. Dados do censo do IBGE mostram que na época das gravações do Projeto Vertentes, a população de Poções era de 44.213 habitantes, a maioria deles residentes na zona urbana do município, salientando que este passou por uma urbanização significativa nos anos 2000, tendo cerca de 71,93% de moradores na sede do município. Dados do IBGE de 2018a mostram um total de 48.861 habitantes.

A economia de Poções gira em torno da agricultura com a produção de café, feijão, tomate, mandioca, banana, milho, aipim e mamona, da pecuária com a criação bovina e há um destaque na mineração com a produção de ferro e amianto. Conforme Silva (2005), o ponto alto das festividades de Poções é a festa do Divino Espírito Santo, padroeiro da cidade e a tradição folclórica do Reisado.

Santo Antônio de Jesus fica localizada a 187km de Salvador e é considerada como a capital do Recôncavo Baiano devido a sua importância no comércio, na indústria e nos serviços da região. Além do destaque econômico, Santo Antônio de Jesus é conhecido por atrair milhares de pessoas durante os festejos juninos, fato que fez o município também ser conhecido por realizar o São João mais popular da Bahia.

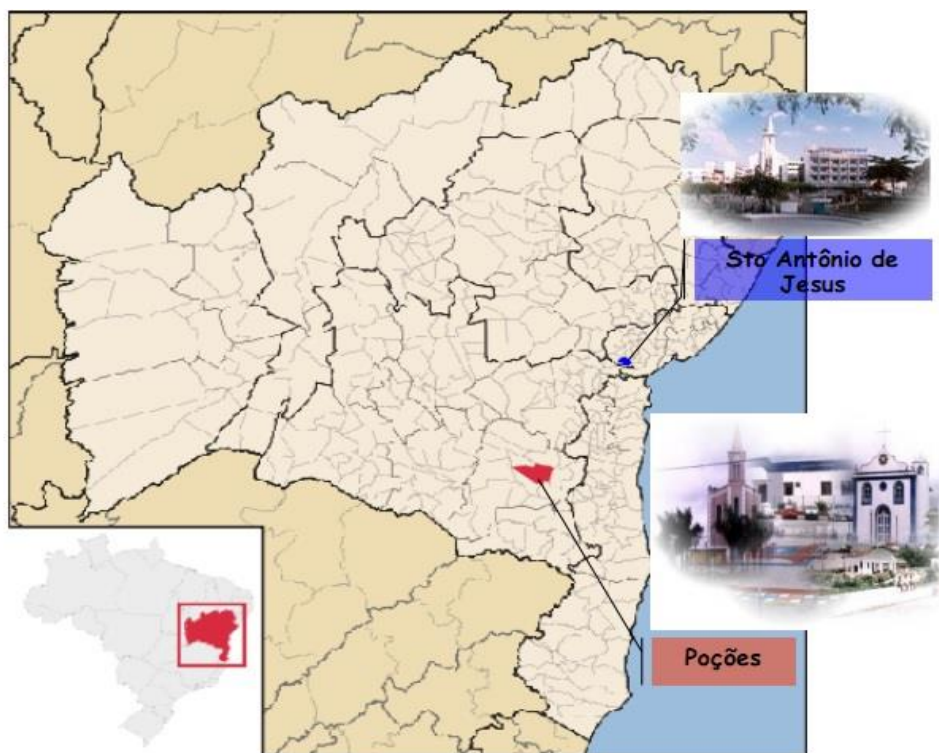
Datam dos séculos XVI e XVII as primeiras expedições na região onde se localiza o município, favorecidas pela passagem do rio Jaguaribe na região. Segundo Araújo e Coutinho (2004), os primeiros habitantes da região eram índios cariris, tupinambás e aimorés, descendentes de Pedra Branca, que a essa época já possuíam, através da Carta Régia, uma proteção e delimitação de reserva de suas terras. Porém, as terras férteis, o cultivo da madeira e os recursos fluviais fizeram com que colonizadores explorassem a região.

No século XVIII, a economia da região era baseada na produção de farinha. Nesse período ocorreram os primeiros assentamentos populacionais. O território de Santo Antônio de Jesus era vinculado à cidade de Nazaré até 1880, quando alcançou independência política e institucional. Dados do IBGE apontam que o município teve um grau elevado de urbanização nos anos 2000. Cerca de 85,62% da população do município residia em sua sede. Na época das gravações das primeiras entrevistas do Projeto Vertentes, ano de 2002, o censo demográfico registrou um total de 77.368 habitantes no município. Os dados do IBGE do censo 2018 apresentam um total de 100.605 habitantes em Santo Antônio de Jesus.

Atualmente, a economia é baseada na agricultura com a produção de amendoim, limão e laranja, na pecuária com a criação de bovinos e muares, na mineração com a produção de areia e argila. Ainda há um grande destaque para o comércio e para o setor de serviços, impulsionados pela migração da população rural para a sede do município. A feira livre da cidade é considerada como a mais barata da Bahia. O município de Santo Antônio de Jesus também é um reconhecido produtor de móveis que atrai diariamente a população das regiões vizinhas, inclusive de Salvador e região metropolitana.

A Figura 1 ilustra o mapa do Estado da Bahia, sinalizando onde se localizam os municípios de Poções e Santo Antônio de Jesus.

Figura 1. Localização das comunidades de Poções e Santo Antônio de Jesus.



Fonte: SANTOS (2016, p. 111).

#### 4.3 ENVELOPE DA VARIAÇÃO

As variáveis linguísticas dependente e independentes que constituem este trabalho foram constituídas partindo do pressuposto de que a concordância consiste

em uma harmonia entre todos os itens constituintes do sintagma (RADFORD, 1997) expressa por uma marca morfológica ou semântica que pode atuar na configuração da concordância variável do SN completo. Diante disso, a análise da CN, neste trabalho, ocorre no nível sintagmático e a variável dependente consiste no SN como um todo.

Foram elencadas doze variáveis independentes, sete linguísticas e cinco sociais, as quais seguiram a chave de codificação elaborada por Lucchesi. As doze variáveis são as seguintes: *configuração sintagmática do SN, realização do núcleo, número de constituintes do SN, número de constituintes flexionáveis do SN, saliência fônica, classe gramatical do constituinte marcado, posição linear do constituinte marcado, sexo, faixa etária, escolaridade, estadia fora da comunidade e localidade*, cujas hipóteses encontram-se descritas no Capítulo 5 juntamente com a análise dos dados com a finalidade de o leitor fazer uma leitura mais dinamizada.

A variável dependente e as variáveis independentes são descritas nas subseções a seguir.

#### 4.3.1 A variável dependente

No SN como um todo, a regra de CN é expressa pela inserção do morfema *-s* em todos os constituintes flexionáveis ou pela presença de forma que semanticamente expressa plural. Porém, como a regra pode ser redundante, nem sempre é aplicada. É possível que nem todos os elementos do SN recebam a marca de plural e, mesmo assim, o sentido da informação não seja prejudicado.

Considerando esses aspectos, a análise sintagmática da CN possui como variável dependente a aplicação ou a não aplicação da regra em todo o SN, sendo atribuído um valor positivo quando há presença de marca de plural em todos os itens flexionáveis do SN, conforme o exemplo (1); e um valor negativo quando não há marca de plural em pelo menos um item flexionável do SN, exemplo (2).

(1) Não, eu tenho **as poucas fitas** [JOSM1]

(2) **os menino** tá zoando muito [ISRAM2]

A partir dos fundamentos empíricos fornecidos por estudos anteriores sobre a CN no português brasileiro (SCHERRE, 1988; ANDRADE 2003, 2006; BAXTER, 2009; BRANDÃO e VIEIRA, 2012, 2013, dentre outros), determinou-se um conjunto de cinco variáveis estruturais para verificar o condicionamento da variável dependente no português popular do interior da Bahia no nível sintagmático. Na subseção a seguir, apresento as variáveis independentes estruturais.

#### 4.3.2 As variáveis estruturais

Para a análise sintagmática da CN neste trabalho, foram estabelecidas sete variáveis estruturais, cujas hipóteses são apresentadas juntamente com os resultados e a análise dos dados no Capítulo 5. Nesta subseção, descrevo a estruturação dos níveis de fatores de cada variável estrutural.

A variável *configuração sintagmática do SN* diz respeito à estrutura do sintagma e busca verificar se um determinado tipo de constituição favorece mais a presença de marcas de plural no SN do que outro (SCHERRE, 1988; ANDRADE, 2006). Devido às diversas possibilidades de combinações de constituintes na formação do SN, inicialmente fixei oito estruturas diversas das estruturas sintagmáticas a fim de não comprometer a análise probabilística dos dados. O Quadro 5 mostra as estruturas propostas, considerando o núcleo e os constituintes adjacentes.

Quadro 5. Fatores da variável *configuração sintagmática do SN*.

Fatores	Exemplos
SN = determinante + nome	pa fazê <b>as malineza</b> <sup>20</sup>
SN = numeral + nome	ela pegô <b>duas esponjona</b>
SN = quantificador indefinido + nome	são <b>todos amigos</b>
SN = nome + quantificador indefinido	<b>os prato tudo</b>
SN = nome + adjetivo	era <b>esses instrumentos mais velhos</b>
SN = adjetivo + nome	inventaro fal... <b>falsas promessa</b> pra mim
SN = pronome possessivo + nome	é coisa <b>dos seus tempo</b>

<sup>20</sup> Expressão da variedade popular do PB que significa *maldade*.

SN = nome + pronome possessivo	<b>amigos meus...</b> aqui
--------------------------------	----------------------------

Fonte: elaborado pela autora.

Na análise qualitativa dos dados, também observei que as ocorrências dos pronomes indefinidos no SN foram palavras que carregam a semântica de quantificação, como mostram os exemplos listados em (3).

- (3)
- a. **certos** anos atrás [JOJM28]
  - b. num tem **muitas** pessoas aqui [MAGF26]
  - c. a televisão veio **pôcos** tempos [ANTM25]
  - d. teve **várias** mulheres [MANM76]
  - e. **ôtas** zonas pra fora aí [VALM46]

Os pronomes indefinidos destacados no exemplo (3) possuem propriedades gramaticais, semânticas e discursivas que nos permitem classifica-los como *quantificadores indefinidos*. Segundo Castilho (2010, p. 486, 505-506), várias classes gramaticais funcionam para a operação de quantificação no português: o morfema -s, o artigo pluralizando o SN, os pronomes e os advérbios quantificadores, e ainda os substantivos coletivos. Castilho (2010) cita o que afirma Mattos e Silva (1989)

Os quantificadores pronominais são atualizadores do nome que não se estabelecem relação entre o que é designado pelo nome e as pessoas envolvidas no processo da comunicação, como os demonstrativos e possessivos, mas que acrescentam ao que é designado alguma informação sobre a quantidade (MATTOS E SILVA, 1989, p. 188)

E explica que a informação sobre quantidade pode ser definida, no caso dos numerais, e indefinida, quando há presença dos quantificadores indefinidos, chamados de pronomes indefinidos pela gramática tradicional (CASTILHO, 2010). Os quantificadores indefinidos possuem um estatuto pronominal e integram uma classe heterogênea, podendo se apresentar na morfossintaxe como núcleo do SN, como pronomes ou como determinantes (cf. CASTILHO, 2010, p. 505-510).

Compreendendo a concordância como um fenômeno que envolve quantificação, como tenho mostrado ao longo desta tese, e entendendo que a “quantificação não é simplesmente uma ‘classe gramatical’” (BAGNO, 2011, p. 829), amalgamei as ocorrências dos pronomes indefinidos da amostra em estudo no âmbito dos quantificadores. Além disso, as ocorrências de SNs com os itens lexicais “os

outros”, exemplificado em (4), foram consideradas como [determinante + nome] devido também as suas propriedades morfossintáticas, semânticas e discursivas.

(4) na rua **dos ôtros** [EDM34]

Devido às várias possibilidades de formação de estruturas de SN no português, foi estabelecida uma hierarquia de acordo com a chave de codificação: nos casos em que mais de um modificador está presente no SN, a codificação obedece a seguinte hierarquia, definida considerando o caráter semântico do constituinte.

1. O quantificador *todas/todos/tudo* é o mais relevante:

Ex: *Meus irmão mais velho **tudo*** = [nome + quantificador]

2. O adjetivo vem em seguida:

Ex: *Algumas **mulé** mais velha* = [nome + adjetivo]

3. Depois, os demais quantificadores indefinidos:

Ex: ***Alguns** filho meus* = [nome + indefinido]

4. Por fim, os pronomes possessivos:

Ex: *Esses **meus** irmão* = [possessivo + nome]

A quantificação preliminar da variável configuração sintagmática do SN em relação à variável dependente revelou falta de ortogonalidade com a variável número de constituintes flexionáveis do SN, ou seja, uma distribuição muito irregular dos dados entre os níveis da variável, principalmente nos seguintes fatores: [nome + pronome possessivo] e [pronome possessivo + nome]. Em virtude da falta de ortogonalidade, amalgamei: os dados de [nome + pronome possessivo] ao grupo [nome + adjetivo]; e [pronome possessivo + nome] ao grupo [adjetivo + nome], levando em consideração as semelhanças distribucionais, sintáticas e semânticas observadas entre esses elementos em relação ao núcleo do SN, reduzindo o número de oito fatores para seis fatores, especificados em (5), cujos resultados são apresentados na subseção 5.2.1.

(5) a. [determinante + nome];



- b. [numeral + nome];
- c. [nome + quantificador indefinido];
- d. [quantificador indefinido + nome];
- e. [nome + adjetivo];
- f. [adjetivo + nome].

Com a análise da variável *realização do núcleo*, pretendo analisar se os sintagmas que possuem núcleo são mais propensos à realização da CN em relação aos sintagmas cujos núcleos não são realizados. Essa variável baseia-se função textual do SN, na qual a estrutura sintagmática pode ter relação com ideias fora do texto, tais como: relação espacial, tempo, medida e valor, o falante tende a marcar o plural para reforçar ao que se refere (SCHERRE, 1988). Dois fatores foram atribuídos a essa variável, conforme o Quadro 6.

Quadro 6. Fatores da variável *realização do núcleo*.

Fatores	Exemplos
SN com núcleo	chegava aqui <b>os meninos</b> brincando
SN sem núcleo	eu também sei respeitá <b>os mais velhos</b> <sup>21</sup>

Fonte: elaborado pela autora.

A análise da variável *número de constituintes do SN* avalia o efeito da quantidade de itens do SN sob o uso de uma ou outra variante da CN, sendo determinados quatro fatores para essa observação. Com o intuito de analisar essa variável, foram estabelecidos os fatores registrados no Quadro 7.

Quadro 7. Fatores da variável *número de constituintes do SN*.

Fatores	Exemplos
Dois	<u>os filhos</u> 1 2 <u>vinte anos</u> 1 2
Três	<u>Uns quatro meses</u> 1 2 3

<sup>21</sup> A palavra **velho**, que pode ocorrer como substantivo e como adjetivo, foi analisada como adjetivo devido à modificação pelo adverbio **mais**.

Quatro ou mais	<u>Os meus irmãos mais velhos</u>			
	1	2	3	4

Fonte: elaborado pela autora.

A variável *número de constituintes flexionáveis do SN* considera a quantidade de itens presentes no SN que podem ser pluralizados e como isso pode condicionar a aplicação da regra de CN. O Quadro 8 apresenta a estruturação dos níveis delimitados para a análise dessa variável.

Quadro 8. Fatores da variável *número de constituintes flexionáveis do SN*.

Fatores	Exemplos
Um constituinte	até <b>quinze dias</b>
Dois constituintes	com <b>meus filhos</b>
Três constituintes	<b>aquelas calça branca</b>
Quatro ou mais constituintes	<b>as ôtras meninas pequena</b>

Fonte: Elaborada pela autora

A variável *saliência fônica* dos elementos do SN, analisada na perspectiva sintagmática, segue a tendência no que tange à diferenciação morfofonológica das palavras. São analisados os SNs em que todos os itens possuam plural regular e os SNs em que, pelo menos um item, apresente diferenciação fônica quando pluralizado. Os níveis de fatores e seus respectivos exemplos estão elencados no Quadro 9.

Quadro 9. Fatores da variável *saliência fônica*.

Fatores	Exemplos
SN com todos os itens regulares	<b>As moçaS</b> iam namorá
SN com pelo menos um item que apresenta diferenciação fônica quando pluralizados (plural irregular)	<b>Os vereadorES</b> , devia elegê um aqui dento

Fonte: elaborado pela autora.

Embora a análise da CN neste trabalho esteja centrada na perspectiva sintagmática, alguns aspectos mórficos necessitam ser considerados. Nesse sentido, estabeleci a *classe gramatical do constituinte marcado* e a *posição linear* desse constituinte como variáveis linguísticas para analisar: (i) quais das classes de palavras

flexionáveis do PB são mais favoráveis a receber a marca de plural, e (ii) qual a posição linear em que mais é frequente o constituinte marcado.

As classes de palavras foram elencadas em: artigos, pronomes possessivos, pronomes demonstrativos, quantificadores indefinidos, substantivos<sup>22</sup> e adjetivos, conforme apresenta o Quadro 10. Porém, na proporção em que a configuração do SN apresentava mais de um elemento marcado, os fatores foram sendo acrescentados, como veremos na Tabela 10 da subseção 5.2.6, no Capítulo 5.

Quadro 10. Fatores da variável *classe gramatical do constituinte*.

Fatores	Exemplos
Artigos	<b>as</b> coisa / <b>umas</b> festa daqui
Pronome demonstrativo	<b>esses</b> dia eu tava brincano
Pronome possessivo	Brincava com <b>meus</b> irmão
Quantificador indefinido	<b>muitas</b> coisa
Adjetivo	Tenho <b>ótimos</b> vizinho

Fonte: elaborado pela autora.

Para a variável posição linear do constituinte marcado foram estabelecidos os seguintes fatores<sup>23</sup>:

Quadro 11. Fatores da variável *posição linear do constituinte marcado*.

Fatores	Exemplos
Primeira posição	<b>As</b> festa
Primeira posição e segunda posição	Por causa <b>das meninas</b> doente

Fonte: elaborado pela autora.

A análise de cada uma das variáveis é feita no Capítulo 5. Na subseção a seguir, apresento as variáveis sociais estabelecidas para este trabalho.

<sup>22</sup>A única ocorrência de substantivos em primeira posição do SN foi: “Amigos meus” [ANTM25]. Tendo em vista que este trabalho não analisou os denominados nomes nus (MÜLLER, 2002), e por ser um único dado, essa ocorrência não foi considerada. Além disso, esse dado não poderia ser analisado na perspectiva mórfica que adotei para averiguar os dados de ausência com as variáveis classe e posição do constituinte marcado (cf. subseção 5.2.6), já que apresenta presença de concordância.

<sup>23</sup> Todos os exemplos elencados nos grupos de fatores das variáveis foram extraídos do *corpus* deste trabalho.

### 4.3.3 As variáveis sociais

Segundo Labov (2008 [1972]), as variáveis sociais estão interligadas ao comportamento linguístico da variação e da mudança. Neste trabalho, estão presentes as variáveis sociais mais utilizadas nos estudos variacionistas brasileiros: *sexo*, *faixa etária* e *escolaridade*. Porém, devido às especificidades das comunidades de fala analisadas, mais duas variáveis sociais foram especialmente consideradas: *estadia fora da comunidade* e *localidade*. A seguir, são detalhadas cada uma delas.

A variável *sexo*, nos estudos do PB, apresentam duas tendências: (i) que as mulheres impulsionam a mudança utilizando a variável de prestígio quando se trata de comunidades urbanas (LUCCHESI, 2009); (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). Dessa forma, esta pesquisa busca saber qual é o efeito do sexo do falante sobre a variável dependente nas comunidades de Poções e Santo Antônio de Jesus.

A análise da variável *faixa etária* leva em consideração as diferentes gerações de informantes que, neste trabalho, estão distribuídas em três faixas etárias - faixa I: 25 a 35 anos; faixa II: 45 a 55 anos; faixa III: mais de 65 anos. Essa variável permite identificar se na comunidade estudada ocorre variação estável ou um processo de mudança em curso, de maneira que associada a outras variáveis, como a escolarização, por exemplo, esse quadro fica ainda mais evidente.

A variável *estadia 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 às comunidades de Poções e Santo Antônio de Jesus, acredito que não somente a urbanização, mas a necessidade de sair das comunidades para outras localidades mais desenvolvidas, inclusive para Salvador e/ou São Paulo como mencionam alguns informantes, condicionem a aplicação da regra de CN no SN.

A *escolaridade* é uma variável bastante relevante nos estudos sociolinguísticos porque é através da instrução formal que a norma linguística de prestígio é mais propagada. De maneira geral, os estudos sobre o PB têm mostrado que a variável

*escolaridade* possui uma determinada tendência: os falantes com maior nível de escolarização tendem a utilizar mais a norma padrão, enquanto a classe intermediária e a classe mais baixa, com menor grau de escolaridade, conservam os traços linguísticos mais populares (SCHERRE, 1988; LOPES, 2001; ANDRADE, 2006).

Para as comunidades de Poções e Santo Antônio de Jesus, suponho que essa variável siga a mesma tendência. Para verificar isso, os informantes foram divididos em dois grupos: analfabetos, aqueles que nunca tiveram contato com a instrução formal; e os semianalfabetos, aqueles que frequentaram a escola por um determinado período, mas não concluíram. Vale ressaltar que nenhum dos 48 informantes cursou o ensino básico completo.

A variável *localidade* considera o local em que residem os falantes levando em conta os processos de urbanização como influenciadores do comportamento linguístico das comunidades no sentido de se incrementar o uso da regra padrão da CN. Essa variável social analisa o comportamento linguístico dos falantes da zona rural e dos falantes da zona urbana dos municípios a fim de verificar diferenças ou não nos dialetos quanto à realização do fenômeno.

#### 4.4 O SUPORTE ESTATÍSTICO COMPUTACIONAL E O TRATAMENTO DOS DADOS

Nesta seção apresento os procedimentos adotados para o tratamento dos dados da presente pesquisa e os principais aspectos do programa computacional utilizado para obtenção dos resultados estatísticos.

A análise estatística que norteou o tratamento dos dados da pesquisa empreendida transcorreu a partir da perspectiva da estatística descritiva e da estatística inferencial. A estatística descritiva, permite que o pesquisador utilize e compreenda medidas de correlação entre duas ou mais variáveis de interesse. A estatística inferencial, por sua vez, caracteriza-se como uma análise criteriosa e científica e da interpretação probabilística dos resultados dos métodos correlacionais (GRIES, 2013; LEVSHINA, 2015).

Os processos da estatística descritiva e da estatística inferencial ocorrem simultaneamente, porém é com a estatística inferencial que o pesquisador observa as hipóteses testáveis, como bem sintetiza Barbosa (2013, p. 53):

O modelo teórico é assim testado a partir de medidas de variáveis ligadas diretamente às hipóteses levantadas, medidas que são avaliadas por métodos estatísticos inferenciais que estabelecem uma ponte entre amostra e população estatísticas. Esse último passo permite, ou a confirmação, normalmente acrescida de uma quantificação das variáveis em escrutínio, ou a refutação, ou ainda o refinamento das hipóteses levantadas.

É justamente a falseabilidade das hipóteses, a possibilidade de colocá-las à prova que imprime o rigor científico de um trabalho (GRIES, 2013; OUSHIRO, 2021). Por exemplo, a seguinte afirmação pode ser falseada: "Os informantes da zona urbana do município de Poções aplicam mais a regra de concordância do que os informantes da zona rural". Essa proposição é que se chama **hipótese alternativa** ou **H1**, nos estudos estatísticos. Em oposição a ela, existe a **hipótese nula** ou **H0**, ou seja, a hipótese nula é formulada com a negação da hipótese alternativa (GRIES, 2013). Utilizando o contexto do mesmo exemplo, a H0 seria: "Os informantes da zona urbana do município de Poções não aplicam mais a regra de concordância do que os informantes da zona rural".

Analisar dados seguindo procedimentos estatísticos à luz da ciência, portanto, requer do pesquisador capacidade de seguir determinadas etapas que vão desde a observação dos fatos de interesse a conclusões baseadas na análise da informação contida em uma amostra da população<sup>24</sup>.

Isso posto, após o planejamento e a observação da amostra, realizei o levantamento exaustivo dos dados, que consiste em procurar e destacar, nas transcrições dos inquéritos, o objeto de estudo a ser analisado, no caso desta pesquisa: todas as ocorrências, de SN com ou sem concordância. Os dados utilizados para esta pesquisa foram coletados em 48 inquéritos de informantes dos municípios de Poções e de Santo Antônio de Jesus: 24 inquéritos de cada município, sendo 12 da zona rural e 12 da zona urbana. Essa estratificação foi feita para verificar se existem diferenças nos dialetos quanto à aplicação da CN, bem como para observar o *continuum* de urbanização (BORTONI-RICARDO, 2005) do fenômeno linguístico em análise.

---

<sup>24</sup> Em Estatística, uma população é o conjunto de elementos ou seres portadores de pelo menos uma característica em comum, cujo comportamento interessa a uma análise. Uma amostra é uma parte ou um subconjunto representativo de uma população, é um conjunto de elementos extraídos da população (COSTA, 2011, p. 20-22). No caso desta pesquisa de doutorado, a população é o português falado no interior da Bahia, e amostra é *corpus* formado pelos informantes dos municípios de Poções e Santo Antônio de Jesus.

No processo de levantamento dos dados, observei os contextos em que as variantes ocorriam com a finalidade de verificar quais os fatores condicionadores da variação. Para cada nível de fatores das variáveis, atribuí um valor representado pelo próprio nome do fator ou por siglas que o representasse. A esse processo, chama-se de processo de codificação. A partir dele, os dados foram importados para o sistema do programa computacional utilizado neste trabalho, o programa R (R Core Team, 2021), com a interface RStudio.

Ao longo desta seção, exponho os conceitos básicos dos comandos do R utilizados na rodada dos dados deste trabalho para que os leitores ainda não familiarizados com a linguagem possam se valer do conteúdo para a devida interpretação dos resultados<sup>25</sup>. Vale salientar que o script<sup>26</sup> completo da rodada dos dados desta pesquisa está disponível como apêndice deste trabalho.

Constituído por uma linguagem de programação estatística e gráfica utilizada para manipulação, análise e visualização de dados, o R<sup>27</sup> é um software livre, gratuito, e de código aberto. Atualmente, é um dos programas de análise estatística mais utilizado nas ciências (CHAMBERS, 2020), por ser rápido, eficaz e por oferecer inúmeras possibilidades aos usuários, inclusive a criação de pacotes com funções customizadas para suas análises.

O primeiro passo para a manipulação dos dados é ter uma visualização deles para verificar a sua distribuição. Isso pode ser feito através da tabulação do conjunto de dados em tabelas e gráficos de frequência e proporção. Neste trabalho, a tabulação inicial dos dados gerais indicou a frequência de 4.977 ocorrências da variável dependente: 845 SNs com concordância, equivalendo a uma proporção de 16,9% dos dados, e 4.132 SNs sem concordância, uma proporção de 83,1% do total dos dados.

Dando continuidade à verificação da distribuição dos dados, foram dados comandos ao R para a tabulação de tabelas e gráficos de frequências e proporções de cada uma das doze variáveis independentes desta pesquisa. Todos esses resultados estão descritos no Capítulo 5. Segundo Oushiro (2017), a visualização da distribuição dos dados permite que o pesquisador verifique se eles estão de acordo com as expectativas e com a teoria, possibilitando a tomada decisões sobre eles. A

---

<sup>25</sup> Para uma introdução ao R para linguistas, ver Gries (2013), Levshina (2015), Lima Jr. *et. al.* (2020) e Oushiro (2017; 2021).

<sup>26</sup> Scripts são escritos no editor de códigos do RStudio. É em um script onde se ordena a execução de uma sequência de comandos escritos previamente (GUERRA *et.all.*, 2020).

<sup>27</sup> R é também o nome da linguagem de programação para computação estatística usada no programa.

tomada de decisões envolve, por exemplo, a possibilidade de amalgamar níveis de fatores cujos dados não apresentam ortogonalidade, isto é, não estão bem distribuídos em relação aos dados e em relação à variável dependente.

Além disso, essas inspeções dos dados permitem que o pesquisador verifique quais testes estatísticos são relevantes para a análise da amostra. Neste trabalho, para constatar se as proporções dos níveis de cada variável independente diferiam significativamente ou não em relação à variável dependente, foi utilizado o teste de qui-quadrado. Esse é um teste de *análise univariada*, que verifica se há correlação entre duas variáveis (GRIES, 2013; LEVSHINA, 2015; LIMA JR. *et al.*, 2020; OUSHIRO, 2017; 2021). Assim, apliquei o teste de qui-quadrado (de Pearson) a cada uma das doze variáveis independentes para avaliar se havia diferenças significativas entre proporções das variantes.

A aplicação do teste de qui-quadrado é feita sobre a Tabela de frequência e somente pode ser aplicado em duas variáveis nominais. O resultado desse teste aponta se há correlação ou não entre as variáveis através do valor-*p*. O valor-*p* é a probabilidade de se ter observado tal distribuição dos dados em caso de a hipótese nula ser verdadeira. As proporções são consideradas significativamente diferentes quando o valor-*p* é abaixo de 0,05<sup>28</sup>. Passemos a um exemplo do teste para devida ilustração.

(6) Pearson's Chi-squared test with Yates' continuity correction  
 $X^2 = 13.925$ , (1),  **$p = 0.0001903$**

Em (6), há o resultado da aplicação do teste de qui-quadrado sobre a variável dependente *ausência de CN X presença de CN* e a variável *realização do núcleo*. O valor-*p*, destacado em negrito (**0.0001903**), indica que a diferença entre as proporções é significativa, pois é um valor abaixo de 0,05. Hipoteticamente, podemos inferir que a *realização do núcleo* é uma variável que apresenta proporções que podem ser analisadas para o fenômeno da CN, seja para o favorecimento ou desfavorecimento da regra.

---

<sup>28</sup>Na comunidade científica, convencionalmente foi estabelecido que o valor-*p* <0,05 é aceitável para avaliar a significância do resultado obtido (Gries, 2013; Levshina, 2015; Oushiro, 2017; Lima Jr. *et al.*, 2020).



Neste trabalho, utilizei o teste de qui-quadrado para verificar a diferença entre as proporções de cada variável independente em relação à variável dependente. Após a verificação do valor-p do teste para cada variável, seguida de uma análise qualitativa, foram reportadas para o modelo de análise multivariada as variáveis com valor-p significativo e que apresentavam características qualitativas relevantes para a análise dos dados. Vale ressaltar que o fato de uma variável ser significativa na análise univariada não implica que ela será significativa na análise multivariada, pois o mecanismo de combinações estatísticas é mais complexo. Os resultados indicaram que apenas uma variável não apresentou diferença entre as proporções no teste de qui-quadrado, a variável *número de constituintes do SN*, cuja discussão é feita na subseção 5.2.2 do Capítulo 5 deste trabalho.

Para melhor descrever o comportamento da variável dependente, que é binária, utilizei o modelo estatístico da regressão logística com efeitos mistos (aplicando a função *glmer*). Conforme Godoy (2019), dados de experimentos ou observações que trazem medidas repetidas por participantes – informantes -, itens ou outros fatores não são adequadamente tratados pelo teste do qui-quadrado porque nesse tipo de teste não é possível inspecionar os efeitos de uma variabilidade aleatória que cada informante<sup>29</sup> pode introduzir na análise. Em virtude disso, nesses casos é razoável fazer uma regressão logística.

Nas palavras de Godoy (2019, n.p), “a regressão logística analisa se a chance de sucesso de um grupo A é diferente da chance de sucesso do grupo B”. Neste estudo, por exemplo, a regressão logística mostra a correlação da presença de concordância com a realização do núcleo no SN, indicando se as chances de realização da regra são diferentes, para mais ou para menos, nos SNs com núcleo realizado e nos SNs sem núcleo realizado.

A regressão logística realiza esse processo com todas as variáveis inseridas no modelo, ou seja, o algoritmo da regressão testa a associação das variáveis independentes com a variável dependente fazendo todas as combinações possíveis

---

<sup>29</sup> Uma das vantagens da utilização do R e seus testes estatísticos para os estudos sociolinguísticos é poder controlar a ação de variáveis aleatórias nos modelos de efeitos mistos realizados através da regressão logística. As variáveis aleatórias são aquelas cujos resultados dependem da influência de outros fatores associados a ela e que o pesquisador não tem como controlar (Gries, 2013; Levshina, 2015; Oushiro, 2021). Neste trabalho, por exemplo, controlei como variável aleatória o *informante* no modelo de efeitos mistos para verificar se comportamento individual de algum falante interferiu nos resultados e verifiquei que estes não foram enviesados pela ação dos informantes.

entre elas. O processo se repete “até que não haja variáveis associadas ao resultado no nível pré-definido de significância estatística” (LEVSHINA, 2015, p. 291).

No caso desta pesquisa, fiz o cruzamento das frequências entre as variáveis independentes para observar a (não)existência de células vazias nos fatores de cada variável e verifiquei quais variáveis não poderiam ser dispostas no mesmo modelo de regressão logística. Células vazias entre os fatores podem interferir diretamente no resultado da regressão, uma vez que ela realiza todas as combinações possíveis e não é exequível combinar fatores que não estejam minimamente ortogonais. Compreendi, pois, que esse procedimento prévio é essencial à análise de multivariáveis, conforme diz Tagliamonte (2012, p. 121).

O cruzamento das frequências revelou uma série de lacunas na distribuição dos dados entre as variáveis *configuração sintagmática do SN* e *número de constituintes flexionáveis*, por exemplo: nos níveis de fatores da *configuração sintagmática do SN*, [adjetivo + nome], [determinante + nome], [nome + adjetivo] e [quantificador + nome] não ocorreu nenhum dado de presença de concordância no nível [um item flexionável] *da variável número de constituintes flexionáveis do SN*, enquanto para esse mesmo nível, a estrutura [numeral + nome] apresentou 340 ocorrências de CN. Dessa maneira, não seria adequado um modelo em que essas variáveis estivessem dispostas. Foi necessária a organização de dois modelos, expostos em (7) e (8) para uma análise estaticamente confiável<sup>30</sup>.

```
(7) modelo1 <- glmer(OCORRENCIA ~ CONFIGURACAO.SN + NUCLEO +
SALIENCIA + SEXO + FAIXA.ETARIA + ESCOLARIDADE + LOCALIDADE +
ESTADIA + (1|INFORMANTE), family = binomial, data = dados)
```

```
(8) modelo2 <- glmer(OCORRENCIA ~ SALIENCIA + NUCLEO +
NUMERO.FLEXIONAVEIS + SEXO + FAIXA.ETARIA + ESCOLARIDADE +
LOCALIDADE + ESTADIA + (1|INFORMANTE), family = binomial, data = dados)
```

Além de ser um modelo estatístico adequado para análises multivariadas, a regressão logística gera um coeficiente linear, o *intercept*<sup>31</sup>, e coeficiente angulares

---

<sup>30</sup> É de conhecimento que um pesquisador não deve realizar diversas análises com o mesmo conjunto de dados a fim de encontrar o resultado desejado. Manipular os dados e as análises está entre as práticas que caracterizam a má conduta científica, denominada como *p-hacking* (NUZZO, 2014; LIMA JR e GARCIA, 2021).

<sup>31</sup> O intercept informa o valor estimado pelo modelo a partir do nível de referência de cada variável. De modo geral, o R lê o nível de referência de acordo com a ordem alfabética ou também, pela ordem de disposição ordinal, o segundo fator de cada variável é posto como o nível de referência. O pesquisador deve estar atento a isso para efetuar a leitura correta dos resultados.

para cada variável independente dentro do modelo estatístico (OUSHIRO, 2017). O modelo avalia se a estatística difere significativamente de zero (valor-p) e gera medidas que são chamadas de *logodds* – logaritmos de chances -. Além dos valores de estimativa em *logodds*, o processamento estatístico dos dados pode gerar valores em *odds-ratio* ou *odds* – chances – e em probabilidade, sendo que todas essas medidas estatísticas possuem relação entre si, são medidas probabilísticas capazes de expressar o resultado de um teste cuja variável dependente tem distribuição binomial. A diferença está em como os valores se apresentam devido às escalas de medidas, cabendo ao pesquisador escolher qual a melhor forma de reportar os resultados de sua pesquisa.

Os valores de *logodds* estimam a probabilidade de ocorrência de um fator em relação ao intercept de ocorrência dos níveis da variável dependente, sua escala vai de menos infinito a mais infinito, com ponto neutro em zero. Segundo Oushiro (2017), um *logodds* de zero, ou próximo a ele, indica prontamente se há diferenças significativas, sendo que valores positivos indicam tendência a favorecimento em relação ao nível de referência da mesma variável independente, e valores negativos indicam tendência a desfavorecimento em relação ao nível de referência da mesma variável independente. Por exemplo, se o resultado de um modelo ocorrer um *logodds* de 0.430 para informantes da zona urbana e um *logodds* de -1.122 para informantes residentes na zona rural, convém indicar que, na fala dos informantes da zona urbana é favorecida a ocorrência de CN.

*Odds* (chances) possui uma escala de vai de zero até mais infinito, cujo ponto neutro é 1. Entende-se que as chances de algo ocorrer são maiores, menores ou iguais às chances de não ocorrer. As chances de algo ocorrer se dão pela divisão entre o número de resultados favoráveis e o número de resultados desfavoráveis, e normalmente são expressas por uma razão (cf. OUSHIRO, 2017; 2021). Segundo Godoy (2019), quando comparamos a chance de sucesso entre dois grupos, o teste estatístico busca responder se a chance de sucesso em um grupo aumenta ou diminuiu em relação ao outro grupo e se essa mudança é significativamente diferente de zero.

Como exemplo, podemos converter os valores de *logodds* citados acima em valores de chances, obtendo o seguinte: 0.430 *logodds* é equivalente a 1.53 chances,

---

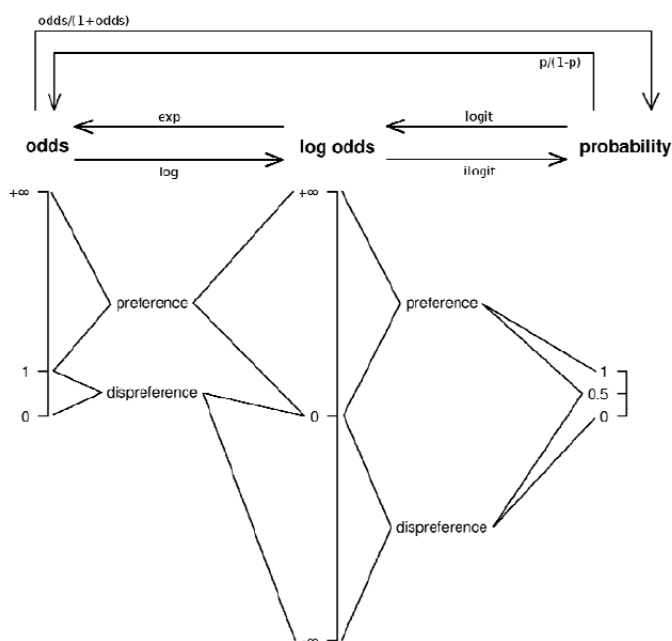
e -1.122 *logodds* equivalente a 0.33 chances. Esse resultado hipotético reporta que: a regra de CN possui chance de 1.53 de ocorrer na fala dos informantes da zona urbana, e 0.32 chances de ocorrer na fala dos informantes da zona rural.

Quanto à medida de probabilidade, o resultado é calculado com a divisão do número total de observações pelo número de resultados favoráveis. A escala de probabilidade é de 0 a 1, em que 0 representa uma chance nula de um evento ocorrer (0%) e 1 representa certeza de um evento ocorrer (100%), sendo o ponto neutro 50% (também reportado na literatura como 0.5). Um resultado que apresente um número abaixo do ponto neutro da escala indica menor probabilidade de que o evento aconteça, e um número acima do ponto neutro indica maior probabilidade de que o evento ocorra (OUSHIRO, 2017; 2021).

Novamente tomando como exemplo os valores citados anteriormente, podemos fazer a conversão para probabilidade e obter: 0.430 *logodds* equivalente a 1.53 chances que corresponde a 65,3% de probabilidade, e -1.122 *logodds* equivalente a 0.332 chances que corresponde a 30,3% de probabilidade. Dessa forma, poderíamos dizer que: a probabilidade de os informantes da zona urbana aplicarem a regra de CN é de 65,3%, enquanto a probabilidade dos informantes da zona rural é de 30,3%.

A Figura 2 (GRIES, 2013, p. 300) ilustra a relação entre as três medidas - *logodds*, *odds* e probabilidade – e mostra as operações que podem ser utilizadas para converter uma medida em outra: *exp*, *log*, *logit* ou *ilogit*. Tal como fiz para exemplificar cada uma das medidas no exemplo de interpretação dos dados: a operação *exp* converteu os valores de *logodds* em *odds*; e a operação *ilogit* foi utilizada para converter os valores de *logodds* em probabilidades. Todas essas operações foram realizadas no R.

Figura 2. Relações entre as medidas de *odds*, *logodds* e probabilidade.



Fonte: GRIES (2013, p. 300)

Para a análise da CN neste trabalho, apresento a distribuição dos dados em tabelas e gráficos de frequências e proporções. Os resultados da regressão logística – estimativas e valor-p – foram reportados também em tabelas, apresentadas posteriormente no Capítulo 5

#### 4.5 SÍNTESE

Neste Capítulo apresentei o enquadramento teórico-metodológico da Sociolinguística (LABOV, 2008 [1972]; 1982; 1994), a qual aborda a língua como heterogênea e condicionada por fatores linguísticos e sociais em uma determinada comunidade de fala, a qual, com o objetivo de elucidar os fundamentos da análise dos dados desta pesquisa.

Indiquei a origem dos dados que compõem o *corpus* deste trabalho, fazendo menção ao Projeto Vertentes da UFBA e sua importância para os estudos linguísticos do PB, especialmente em relação às variedades baianas, apresentando também os

principais aspectos sócio-históricos das comunidades em estudo: os municípios de Poções e Santo Antônio de Jesus, localizados no interior do estado.

Fiz a descrição das variáveis dependente e independentes (estruturais e sociais) consideradas para a análise sintagmática da CN neste trabalho, bem como apresentei o tratamento dos dados e o suporte computacional do programa R, utilizado para a análise probabilística e estatística. As hipóteses juntamente com a análise dos resultados são apresentadas Capítulo 5.

## 5. ANÁLISE DOS DADOS

Após a exposição do objeto de estudo no Capítulo 2, bem como a apresentação dos aspectos teóricos e metodológicos que norteiam esta pesquisa, Capítulos 3 e 4, neste Capítulo são expostos os resultados da análise sintagmática da CN. Resultados obtidos através da quantificação dos dados levantados nos *corpora* das comunidades de fala de Poções e de Santo Antônio de Jesus.

A hipótese geral para essa análise é a de que há mais dados de SNs sem concordância que SNs com concordância por se tratar de uma amostra de dados de fala vernácula de comunidades de fala interioranas, tendo em vista a tendência que seguem os trabalhos de fala sobre o PB popular (ANDRADE, 2003, 2006; BAXTER, 2009; LUCCHESI, 2009).

Os resultados das variáveis estruturais e sociais, obtidos através da análise estatística e probabilística no R, são comentados neste capítulo. Na análise das variáveis, teço comentários comparativos com os estudos, principalmente, de Scherre (1988)<sup>32</sup> com falantes do Rio de Janeiro, Andrade (2003, 2006), Baxter (2009) com falantes afro-brasileiros da Bahia, e Lopes (2001) com falantes de Salvador, por fornecerem matrizes teórico-metodológicas pertinentes acerca do fenômeno linguístico em estudo.

### 5.1 RESULTADOS GERAIS

A partir do levantamento exaustivo e da codificação de 48 inquéritos de informantes residentes nos municípios de Poções e Santo Antônio de Jesus, fiz a análise sintagmática da CN, considerando todo o SN como uma unidade de estudo. A variável dependente foi estruturada em: presença de marca de plural em todos os

---

<sup>32</sup> Ressalto que a utilização do estudo de Scherre (1988) foi importante para este trabalho por se tratar de um estudo clássico sobre a CN no PB que apontou variáveis linguísticas importantes para o desenvolvimento de pesquisas posteriores acerca da mesma temática. A comparação de alguns aspectos do estudo de Scherre (1988) com estudos mais recentes, inclusive com os resultados desta pesquisa, são necessários para a verificação do comportamento das variáveis no sentido de se estas continuam atuando de modo semelhante ou não e, se esse comportamento está mais associado a fatores linguísticos ou a fatores sociais que variam de acordo com a comunidade de fala.

itens flexionáveis do SN *versus* ausência de marca de plural em pelo menos um item flexionável do SN, conforme os exemplos (1) e (2), respectivamente.

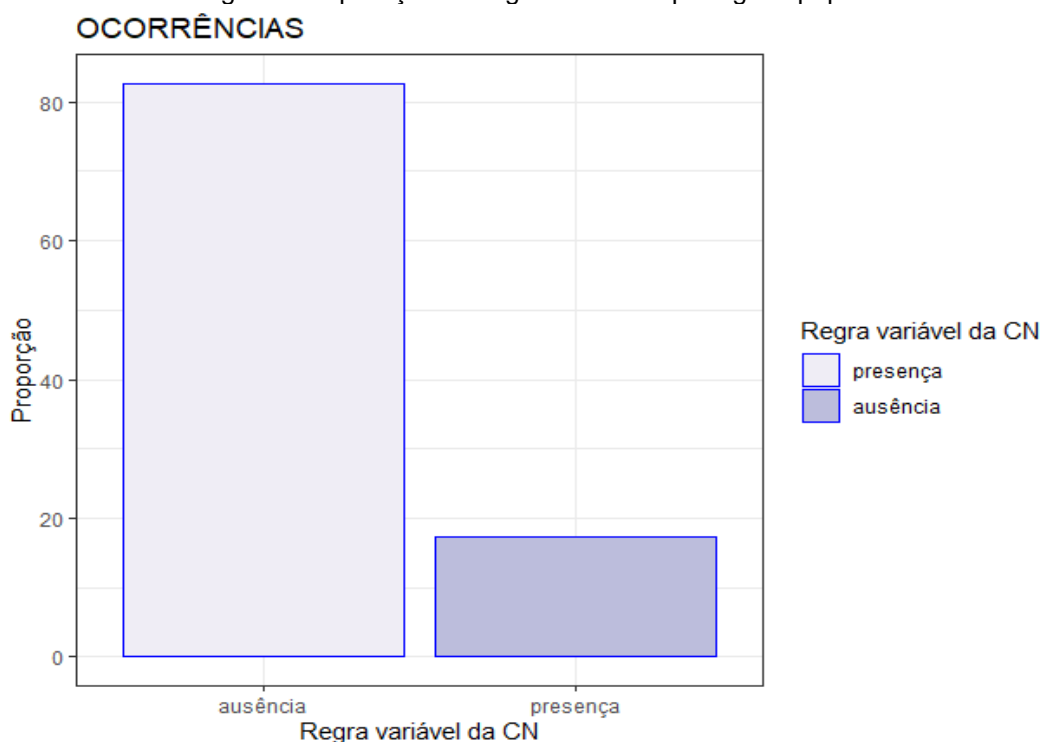
Considerou-se a realização plena da CN, como no exemplo (1), em todo o sintagma a variante inovadora porque contrasta com a variedade popular do PB falada na comunidade, considerando o que diz Lucchesi (2009) sobre a ideia da perda de morfologia flexional no PB popular ser um indicador da influência do contato linguístico. Logo, se houve essa perda, o que é conservador, em comunidades cujas características do contato tendem a ser mais evidentes, é a falta de concordância, como no exemplo (2).

(1) já pedi meu patrão umas três vezes [ANTONM1]

(2) ôtas coisa nunca fui de fazê [JOSEM1]

O processamento inicial dos dados levou à avaliação da distribuição das informações e revelou um total de 4.977 ocorrências da variável dependente. Desse total, foram encontrados 845 SNs com concordância, equivalendo a 16,9% dos dados, e 4.132 sem concordância, cerca de 83,1%. O Gráfico 2 apresenta a proporção desse resultado geral.

Gráfico 2. Resultado geral da aplicação da regra de CN no português popular do interior da Bahia.



Fonte: Elaborada pela autora.



Os resultados gerais da variável dependente mostram que é muito baixo o uso da regra de concordância no português popular do interior da Bahia. Além disso, já podemos perceber que o *status* da CN no *continuum* do português baiano é de quanto mais próxima da capital do estado for a comunidade de fala, maior é aplicação da regra padrão, tendo em vista que Andrade (2006) verificou 8% de presença de CN no português afro-baiano e Lopes (2001), 34% no português de Salvador.

Nas Tabelas 1 e 2, estão postos os resultados gerais dos dois modelos de análise multivariada utilizados para a interpretação dos resultados desta pesquisa. Os valores reportados nas colunas das Tabelas correspondem, para cada nível de fator de cada variável, a: frequência da presença de CN/total das ocorrências de ausência e presença; proporção de aplicação da regra; estimativa em medidas de *logodds*; o erro padrão, que indica a medida de dispersão dos dados na amostra; o valor-z, que indica se os desvios-padrão estão acima ou abaixo da média do valor do erro padrão; e o valor-p, indicando se o resultado é estatisticamente diferente de zero e, portanto, significativo.

Para a análise linguística a que se propõe esta pesquisa, interessa compreender as frequências, as proporções, as estimativas e o valor-p de cada fator. Vale ressaltar que utilizei o modelo 1 para apresentar os dados referentes à variável configuração sintagmática do SN, e as demais variáveis foram analisadas a partir dos valores do modelo 2.

Tabela 1. Estimativas do modelo 1 (de regressão logística, modelo linear generalizado de efeitos mistos) para a CN no português popular da Bahia.

Apl./total = **843/4.867**

Intercept = -1.277 (valor de ref.)

Variável	Apl./Ocorrências	Proporção	Estimativa	Erro padrão	Valor-z	Valor-p
<b>CONFIGURAÇÃO DO SN</b>						
numeral + nome	416/1.617	25,7%	1.031	0.185	5.569	<0.001***
quantificador indefinido + nome	92/404	22,7%	0.648	0.213	3.030	0.002**
adjetivo + nome (valor de ref.)	55/358	15,3%	-1.277	0.441	-2.894	0.003**
determinante + nome	268/2.271	11,8%	-0.010	0.183	-0.056	0.955
nome + adjetivo	12/217	5,5%	-1.161	0.352	-3.292	0.000***
<b>REALIZAÇÃO DO NÚCLEO</b>						
SN com núcleo (valor de ref.)	835/4.832	17,3%	-1.277	0.441	-2.894	0.003**
SN sem núcleo	8/35	22,9%	0.501	0.462	1.084	0.278
<b>SALIÊNCIA FÔNICA</b>						
com saliência (valor de ref.)	261/753	34,7%	-1.277	0.441	-2.894	0.003**
sem saliência	582/4.114	14,1%	-1.026	0.101	-10.153	<0.001***
<b>SEXO</b>						
Feminino (valor de ref.)	506/2.470	20,5%	-1.277	0.441	-2.894	0.003**
Masculino	337/2.397	14,1%	-0.675	0.225	-2.993	0.002**
<b>FAIXA ETÁRIA</b>						
jovem	406/1.451	28%	0.960	0.331	2.900	0.003**
intermediário	242/1.714	14,1%	-0.019	0.280	-0.068	0.945
idoso (valor de ref.)	195/1.702	11,5%	-1.277	0.441	-2.894	0.003**
<b>ESCOLARIDADE</b>						
pouca escolaridade (valor de ref.)	636/2.868	22,2%	-1.277	0.441	-2.894	0.003**
nenhuma escolaridade	207/1.999	10,4%	-0.576	0.302	-1.902	0.057 .
<b>ESTADIA</b>						
viveu fora	415/2.222	18,7%	0.114	0.236	0.484	0.628
nunca viveu fora (valor de ref.)	428/2.645	16,2%	-1.277	0.441	-2.894	0.003**
<b>LOCALIDADE</b>						
zona urbana	498/2.478	20,1%	0.230	0.200	1.150	0.249
zona rural (valor de ref.)	345/2.389	14,4%	-1.277	0.441	-2.894	0.003**

MODELO 1 (OCORRENCIA ~ CONFIGURACAO.SN + NUCLEO + SALIENCIA + SEXO + FAIXA.ETARIA + ESCOLARIDADE + LOCALIDADE + ESTADIA + (1 | INFORMANTE)

Fonte: elaborada pela autora.

Tabela 2. Estimativas do modelo 2 (de regressão logística, modelo linear generalizado de efeitos mistos) para a CN no português popular da Bahia.

Apl./total = **843/4.867**

Intercept = -1.013 (valor de ref.)

Variável	Apl./ Ocorrên- cias	Proporção	Estimativa	Erro padrão	Valor-z	Valor-p
<b>NÚMERO DE CONST. FLEX.</b>						
um	340/1.340	25,4%	0.738	0.092	8.010	<0.001***
dois (valor de ref.)	473/3.187	14,8%	-1.013	0.402	-2.519	0.011*
três ou mais	30/340	8,8%	-0.903	0.211	-4.271	<0.001***
<b>REALIZAÇÃO DO NÚCLEO</b>						
SN com núcleo (valor de ref.)	835/4.832	17,3%	-1.013	0.402	-2.519	0.011*
SN sem núcleo	8/35	22,9%	0.300	0.436	0.436	0.491
<b>SALIÊNCIA FÔNICA</b>						
com saliência (valor de ref.)	261/753	34,7%	-1.013	0.402	-2.519	0.011*
sem saliência	582/4.114	14,1%	-1.076	0.100	-10.746	<0.001***
<b>SEXO</b>						
feminino (valor de ref.)	506/2.470	20,5%	-1.013	0.402	-2.519	0.011*
masculino	337/2.397	14,1%	-0.651	0.222	-2.927	0.003**
<b>FAIXA ETÁRIA</b>						
jovem	406/1.451	28%	1.013	0.327	3.095	0.001**
intermediário	242/1.714	14,1%	-0.009	0.277	-0.034	0.972
idoso (valor de ref.)	195/1.702	11,5%	-1.013	0.402	-2.519	0.011*
<b>ESCOLARIDADE</b>						
pouca escolaridade (valor de ref.)	636/2.868	22,2%	-1.013	0.402	-2.519	0.011*
nenhuma escolaridade	207/1.999	10,4%	-0.559	0.299	-1.871	0.061.
<b>ESTADIA</b>						
viveu fora	415/2.222	18,7%	0.103	0.233	0.446	0.665
nunca viveu fora (valor de ref.)	428/2.645	16,2%	-1.013	0.402	-2.519	0.011*
<b>LOCALIDADE</b>						
zona urbana	498/2.478	20,1%	0.206	0.198	1.040	0.298
zona rural (valor de ref.)	345/2.389	14,4%	-1.013	0.402	-2.519	0.011*

MODELO 2 (OCORRENCIA ~ SALIENCIA + NUCLEO + NUMERO.FLEXIONAVEIS + SEXO + FAIXA.ETARIA + ESCOLARIDADE + LOCALIDADE + ESTADIA + (1 | INFORMANTE)

Fonte: elaborada pela autora.

É importante mencionar que testei modelos de efeitos fixos sem a presença do informante para observar se havia alguma diferença entre os resultados dos efeitos mistos com a presença do informante. Não foram encontradas diferenças quanto à significância das variáveis entre os modelos de efeitos fixos e de efeitos mistos, ou seja, em relação à CN no português popular da Bahia, o comportamento dos informantes não enviesou os dados.

Os resultados da análise univariada do teste de qui-quadrado e da análise multivariada dos dois modelos de regressão logística mostrados nas Tabelas 1 e 2, bem como a análise qualitativa das variáveis são discutidos na seção seguinte.

## 5.2 O ENCAIXAMENTO LINGUÍSTICO DA CN

A análise do encaixamento linguístico da CN na perspectiva sintagmática considerou oito variáveis: *configuração sintagmática do SN*, *realização do núcleo*, *número de constituintes do SN*, *número de constituintes flexionáveis do SN*, *saliência fônica*, *posição da marca de plural* e *classe da posição marcada*.

Nas seções seguintes são comentados os resultados probabilísticos dessas variáveis, considerando as hipóteses levantadas. Ressalto que as variáveis *posição da marca de plural* e *classe da posição marcada* são variáveis que descrevem apenas a frequência e a proporção dos dados de ausência de concordância. Considero necessária a inserção dessas variáveis para verificar e descrever os dados, em quantidade expressiva, de ausência do elemento mórfico marcador de plural -s.

### 5.2.1 Variável *configuração sintagmática do SN*

A variável configuração sintagmática do SN diz respeito à estrutura do sintagma, de acordo com Scherre (1988), na variedade popular do PB, o SN segue a tendência: um determinante e um núcleo. De acordo com a mesma autora, a hipótese para essa variável é de que as estruturas “determinante + nome” e “numeral + nome” favoreceriam a aplicação da CN, enquanto as demais estruturas não seriam favoráveis na mesma proporção.

Lucchesi e Dália (no prelo) afirmam que a correlação hipotética entre a estruturação do conjunto de constituintes do SN e a aplicação ou não da regra de CN gira em torno da simplicidade da estrutura, da adjacência ao núcleo e da redundância da ideia de plural considerando a presença de elementos semanticamente plurais.

Tendo em vista as proposições de Scherre (1988) e de Lucchesi e Dália (2020), no que tange ao favorecimento da regra de aplicação da CN, a hipótese para a variável configuração sintagmática do SN neste trabalho é de que a CN seria mais empregada em SNs com estruturas mais simples e que estruturas que possuam constituintes que reforçam a ideia de plural, como numerais e quantificadores. Vale salientar que o uso do termo “simples” está relacionado com a quantidade de constituintes no sintagma e sua estrutura morfossintática.

A Tabela 3 mostra resultados iniciais sugestivos das frequências e das proporções da configuração sintagmática do SN em relação à aplicação da regra de CN no português popular do interior da Bahia. Podemos ver, na Tabela 3, que as estruturas de SN nas quais a frequência e a proporção de aplicação da CN foram maiores, a saber: [numeral + nome], em que há a presença de um numeral em posição pré-nuclear, conforme o exemplo (3); e [quantificador indefinido + nome], estrutura em que há um quantificador como constituinte anterior ao nome núcleo, como podemos ver em (4) e (5).

- (3) eu já cansei de saí doze horas [ANTM25]  
 (4) a televisão veio pôcos tempos [ANFM25]  
 (5) Com todos os vizinhos, com todos [JOJM28]

Tabela 3. Rodada 1. Frequência e proporção da variável *configuração sintagmática do SN* sobre a aplicação da CN no português popular do interior da Bahia.

<b>Configuração sintagmática do SN</b>	<b>Frequência/Total</b>	<b>Proporção</b>
numeral + nome	416/1.617	25,7%
quantificador indefinido + nome	92/404	22,7%
adjetivo + nome	55/358	15,4%

determinante + nome	243/2.201	11%
nome + adjetivo	12/217	5,5%
nome + quantificador indefinido	2/110	1,8%
<b>TOTAL</b>	<b>845/4.977</b>	<b>16,9%</b>

Fonte: elaborada pela autora.

A tabulação evidenciou que, apesar de amalgamar dois fatores, a distribuição de dados continuou não ortogonal aos fatores da variável número de constituintes flexionáveis do SN, principalmente, no nível do fator [nome + quantificador indefinido]. Essa estrutura de SN apresentou um resultado quase categórico, 1,8% de presença de concordância em oposição a 98,2% de ausência.

Ao fazer uma análise qualitativa dos dados dessa estrutura de SN, observei que, das 110 ocorrências nesse nível, 17 se referem ao quantificador *todo(a)/todos(as)*, exemplos (6) e (7), e a dois únicos dados de presença de concordância são referentes a esse quantificador também, (8) e (9). As outras 93 ocorrências se referem ao quantificador *tudo*, utilizado de forma característica à variedade popular do português, com valor de *todos/todas* e com uma outra particularidade, à qual dediquei a subseção 5.2.2.1 para uma breve descrição.

- (6) E aquelas mania toda [FRAF76]
- (7) tem essas coisa toda [ANTM51]
- (8) As meninas todas são boas [ANTF47]
- (9) passa p'esses órgos todos [MARF45]

Considerando o que diz Labov (2003, p. 243) sobre ser semicategórica a regra linguística que opera entre 95% a 99% de ocorrência, retirei a estrutura [nome + quantificador] por apresentar 98,2% de ausência de CN nas ocorrências dessa estrutura. Vale lembrar que Lopes (2001) analisou a coexistência do quantificador *tudo* com valor de *todos/todas* no estudo da CN no português de Salvador, como já descrito na subseção 2.2.2 do capítulo 2 desta tese, e verificou que o quantificador *tudo* teve correlação com a não aplicação da CN na fala dos informantes descendentes de africanos e aponta que isso pode estar ligado à aquisição linguística

em situação de contato em que, em alguma fase da aquisição, esse quantificador fora interpretado como item que marcava o plural em todo sintagma (cf. LOPES, 2001, p. 342-343).

Os resultados da configuração sintagmática do SN permaneceram idênticos aos que foram mencionados na Tabela 3, com as estruturas [numeral + nome] e [quantificador + nome] com maiores proporções quanto à aplicação da CN: 25,7% e 22,7%, como podemos ver na Tabela 4.

Tabela 4. Rodada 2. Frequências e proporções da variável *configuração sintagmática do SN* sobre a aplicação da CN.

<b>Configuração sintagmática do SN</b>	<b>Frequência/Total</b>	<b>Proporção</b>
numeral + nome	416/1.617	25,7%
quantificador indefinido + nome	92/404	22,7%
adjetivo + nome	55/358	15,4%
determinante + nome	268/2.271	15,3%
nome + adjetivo	12/217	5,5%
<b>TOTAL</b>	<b>843/4.867</b>	<b>17,3%</b>

Fonte: elaborada pela autora.

Na análise univariada com o teste de qui-quadrado, o valor-p < 0,001, ou seja, abaixo de 0,05, indicou que as proporções da variável configuração sintagmática do SN foram significativamente diferentes entre si. Dessa forma, inseri a variável independente em análise na regressão logística com efeitos mistos.

O modelo estatístico da regressão logística, conforme a Tabela 1, apontou que essa variável não obteve uma total correlação com os resultados da variável dependente, quando combinados com as outras variáveis independentes em estudo. Embora não tenha se mostrado estatisticamente relevante na análise multivariada, a variável configuração sintagmática do SN é importante para a análise da CN por indicar resultados sugestivos quanto à presença de constituintes semanticamente plurais favorecer a aplicação da regra de concordância.

Para a variável configuração sintagmática do SN, os resultados indicaram que, há um favoecimento à aplicação da CN nas estruturas [numeral + nome] (1.031 *logodds*) e [quantificador indefinido + nome] (0.648 *logodds*), respectivamente, como mostram os exemplos (10) e (11).

(10) fiquei pagando nove meses [MARF75]

(11) brincava com muitos brinquedos [VALM46]

As demais estruturas elencadas nessa variável não favoreceram a aplicação da CN: [determinante + nome] -0.010; [nome + adjetivo] -1.161 e [adjetivo + nome] -1.277 *logodds*. Os exemplos de (12) a (14) correspondem a essas estruturas, na mesma ordem.

(12) **as meninas** são muito criativas [MAGF26]

(13) era **esses instrumentos mais velhos** [ANTM65]

(14) tem ôtos **novos amigos** [JOSM28]

No estudo de Scherre (1988), configuração do SN [quantificador + artigo definido + nome], que corresponde, neste trabalho, a [quantificador indefinido + nome] foi a estrutura que obteve maior percentual de aplicação da CN, 82% (SCHERRE, 1988, p. 304). Semelhantemente, o estudo de Andrade (2006) acerca das comunidades afro-brasileiras baianas, a estrutura [quantificador + nome] foi a que mais favoreceu a CN, 35% de aplicação da regra.

Os resultados probabilísticos quanto à estrutura do SN na CN no português popular do interior da Bahia indicam que contextos semanticamente pluralizados, com a presença de um numeral ou quantificador indefinido, favorecem mais a presença de concordância do que as estruturas cujos constituintes possuem morfemas de plural mais abstrato, como é o caso das estruturas formadas por [determinante + nome]. Dessa forma, a hipótese de que estruturas mais simples favorecem a concordância no SN não se confirma neste trabalho. No entanto, em relação às estruturas com itens de transparência semântica de plural, a hipótese se confirma. Em outras palavras, a aplicação da regra de CN no português popular do interior da Bahia possui mais associação com o contexto de plural do que com a simplicidade da estrutura.



### 5.2.1.1 O quantificador tudo

O levantamento dos dados, cujos SNs possuem em sua estrutura o quantificador tudo, nos levou a uma observação interessante: o uso do *tudo* em algumas sentenças tem interpretação ambígua, pode ser interpretado como quantificador, tendo como escopo o núcleo do SN ou como intensificador tendo como escopo um adjetivo. A sentença em (15), retirada do *corpus* deste trabalho, é um exemplo desse uso.

(15) As pessoas ***tudo*** legais. [GUISF56]

A ocorrência em (15) pode evidenciar que *tudo* se comporta como *todas/todos*, mas também pode assumir o papel de um advérbio de intensidade, conforme os exemplos em (16) e (17) demonstram, respectivamente.

(16) As pessoas ***todas*** legais.

(17) As pessoas ***muito*** legais.

Essa observação se faz ainda mais pertinente dado o fato de que, mais adiante no inquérito do mesmo informante e no mesmo contexto de conversação, temos a ocorrência (18) que pode indicar uma variação no uso do *tudo* na variedade popular do português do interior da Bahia. O informante faz uso do advérbio de intensidade *muito* e, em seguida, usa o *tudo* na construção sintática da mesma sentença.

(18) INF: Então, eu tenho os vizinho, num tenho o que dizê, poque cada um ne suas casa, “Bom dia”, “Boa tarde”, é assim. **Uns pessoal *muito* legal.**

DOC: Num liga som alto?

INF: Não.

DOC: Humhum.

INF: **Uns pessoal *tudo* legais, aqui, *tudo... tudo.*** [GUISF56]

Godoi (2005) e Godoi e Cançado (2006) analisaram o *tudo* como quantificador no português a partir de duas possibilidades: o uso peculiar do *tudo* no português

popular (19); e o uso “tradicional” do *tudo*, que pode ser traduzido pela expressão *todas as coisas* (20).

- (19) Os menino *tudo* leram o livro do Harry Potter.
- (20) a. Os meninos comeram *tudo* que havia no prato.  
b. Os meninos comeram *todas as coisas* que havia no prato.

(GODOI e CANÇADO, 2006, p. 157)

Apesar de realizarem um trabalho descritivo voltado à análise interna do quantificador, as autoras acreditam que esse fenômeno com o uso do *tudo* se trata de variação, quiçá uma mudança no sistema de quantificadores no PB (GODOI e CANÇADO, 2006). A constatação peculiar sobre a qual debruçaram investigação linguística se percebe que a paráfrase do *tudo* coloquial com a expressão *todas as coisas* não funciona, conforme podemos ver em (21). A melhor paráfrase para esse uso peculiar da palavra *tudo* parece ser a palavra *todos*, expressa em (22).

- (21) a. Os menino *tudo* leram o livro do Harry Potter.  
b. \* Os menino *todas as coisas* leram o livro do Harry Potter.
- (22) a. Os menino *tudo* leram o livro do Harry Potter.  
b. Os meninos *todos* leram o livro do Harry Potter.

(GODOI e CANÇADO, 2006, p. 157, 158)

Trannin (2016, 2017) estudou a sintaxe do quantificador *tudo* no PB popular a partir de uma perspectiva diacrônica, investigando a origem da variação *todo(s)/tudo*. Com o aporte teórico da gramática gerativa, Trannin descreveu a sintaxe e a semântica dos dados e verificou que há duas gramáticas de *tudo*, que diferem quanto à propriedade de seleção do quantificador e sua relação com o associado nominal, e comparou as variedades do PB popular e do português afro-brasileiro. A autora sugere que a sintaxe de *tudo* no PB é semelhante no português afro-brasileiro, que também apresenta a variação entre os quantificadores.

Assim, propus duas gramáticas para o elemento *tudo*: a gramática restritiva estaria presente no PBCF, enquanto a gramática radical seria encontrada no PBP. Nos dados das comunidades quilombolas em estudo é possível observar uma distinção quanto à realização de *tudo* nos contextos sintáticos

de uso. Aparentemente, nas comunidades de Milho Verde (MG) e de Bom Despacho (MG), tudo é usado como variante do quantificador todo como intensificador e em contextos de flutuação. Já na fala dos moradores de Jurussaca (PA) e em Furnas de Boa Sorte (MS), tudo ocorre nos mesmos contextos de todo. Pode-se dizer, portanto, que nas comunidades Bom Despacho e Milho Verde, a gramática de tudo se aproxima da gramática restritiva, enquanto Jurussaca e Furnas de Boa Sorte apresentam a gramática radical. A presença desse fenômeno nas variedades africanas do português e as semelhanças com o crioulo do Cabo Verde favorecem essa hipótese e corroboram com a ideia da existência de um continuum afro-brasileiro do português (TRANNIN, 2016, p. 157-158).

Tendo em vista a relação da concordância com os quantificadores, nesse caso, uma possível relação da falta de concordância com a presença do *tudo*, e o seu uso variando ora como quantificador ora como intensificador, destaco que esse comportamento ambíguo do *tudo* pode justificar o comportamento da CN no *corpus* em análise neste trabalho.

Os dados deste trabalho indicam que, quando há a presença do quantificador *tudo* com valor de *todos/todas* a não aplicação da regra é praticamente categórica, 98,2% de ausência de concordância, conforme mostrado na Tabela 3. Esse resultado indica que o *tudo* pode ser um pluralizador, como sugeriu Lopes (2001), que parece estar associado a alguma fase da aquisição do português em que os falantes reanalisaram o *tudo* como como quantificador geral (cf. LOPES, 2001, p. 342-343).

Ademais, considero esse tema um caminho profícuo às pesquisas linguísticas e corroboro a necessidade de investigar o uso do *tudo* no interior da Bahia para ampliar essa discussão sobre o uso dos quantificadores no PB e os efeitos do contato linguístico.

### 5.2.2 Variável *número de constituintes do SN*

A expectativa de análise para a variável *número de constituintes do SN* é a de que a CN seja favorecida nos sintagmas cuja estrutura possua menos constituintes, se considerarmos o princípio da simplicidade<sup>33</sup>, segundo Lucchesi (2000). Em contrapartida, os SNs com uma quantidade maior de constituintes, devido à estrutura

---

<sup>33</sup> Isso se deve ao pressuposto de que, na fala popular, a ocorrência de sintagmas considerados simples é mais comum, ressaltando o que Lucchesi (2000) chamou de princípio da simplicidade.

mais complexa, há uma tendência à não marcação de plural em todos os constituintes (SCHERRE, 1988; LOPES, 2001).

Em função do baixo número de ocorrências, os fatores SNs com quatro e cinco ou mais constituintes foram amalgamados. As frequências e os percentuais de aplicação da regra encontram-se na Tabela 5.

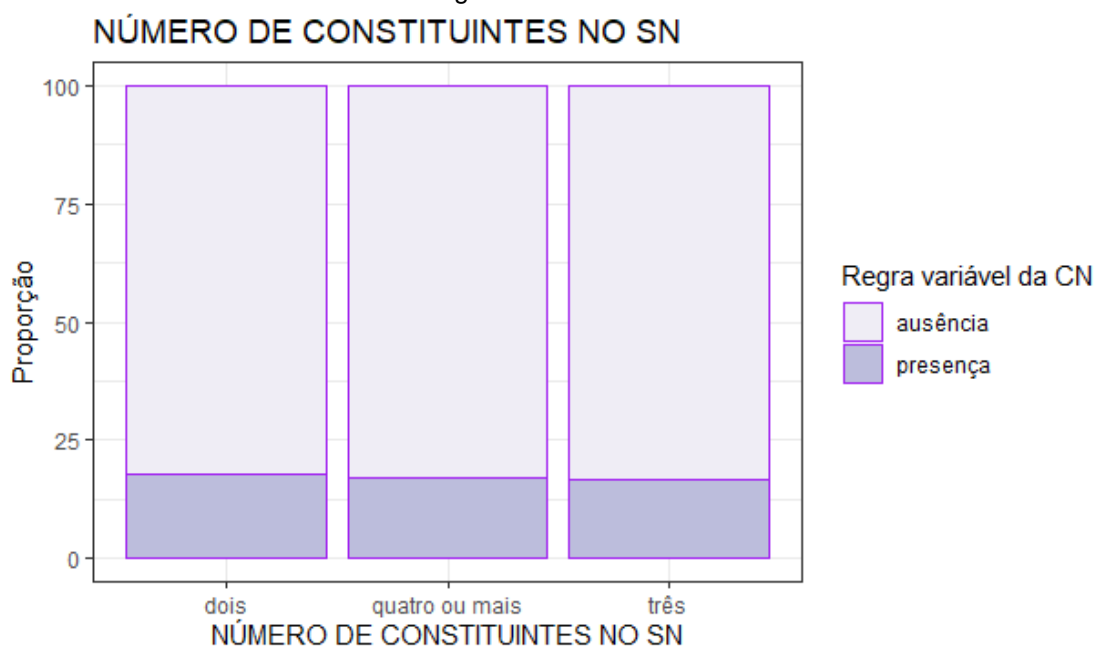
Tabela 5. Frequências e proporções gerais da variável *número de constituintes do SN*.

<b>Número de constituintes do SN</b>	<b>Frequência/Total</b>	<b>Proporção</b>
Dois	530/3.000	17,7%
Três	269/1.610	16,7%
Quatro ou mais	44/257	17,1%
<b>TOTAL</b>	<b>843/4.867</b>	<b>17,3%</b>

Fonte: Elaborada pela autora.

Os resultados mostram uma proximidade na proporção de aplicação da CN independentemente da quantidade de constituintes do SN. Para melhor visualização dos resultados, segue o Gráfico 3 referente às proporções.

Gráfico 3. Proporções da variável número de constituintes do SN em relação à aplicação da regra de CN.



Fonte: elaborado pela autora.

A visualização do Gráfico 3 permite verificar que a diferença entre as proporções dos fatores dessa variável é bastante pequena. O resultado aproximado entre as proporções dessa variável levou a análise univariada com o teste do qui-quadrado a gerar um valor-p 0.711, ou seja, maior que 0,05, confirmando que não há diferença significativa quanto à distribuição dos dados para a variável número de constituintes em relação à aplicação da regra de concordância nos SNs. Em virtude disso, essa variável não foi incluída na análise multivariada da regressão logística.

Por outro lado, os resultados dessa variável podem ser sugestivos quanto à estruturação geral dos SNs no português popular. Os dados parecem apontar que, na fala coloquial, predominam os SNs com dois constituintes, compostos geralmente por um determinante e um nome. No entanto, a regra de CN parece ser mais aplicada nos SNs em que um dos constituintes apresenta marca semântica de plural, como vimos na subseção 5.2.1 sobre a variável configuração sintagmática do SN.

Nas pesquisas de Scherre (1988), Lopes (2001) e de Andrade (2006), os resultados se assemelham aos encontrados nesta análise: quanto menos constituintes tiver o SN, maior a possibilidade de haver CN. As referidas autoras apontam que isso ocorre porque a estrutura do SN no PB popular tende a ser simplificada. Lucchesi e Dália (no prelo) ainda reforçam que

A frequência de SNs mais complexos, com um número maior de constituintes, seriam próprios da fala formal e sobretudo da escrita; sendo, obviamente, uma possibilidade contida na competência linguística do falante, cuja potencialidade pode ser ampliada em condições especiais de uso da língua. Todavia, na fala vernacular, estes SNs desfavorecem a aplicação da regra de concordância, seja porque sobrecarregam o processamento da fala com um grande número de constituintes a serem marcados, seja porque envolvem estruturas menos coesas com adjunção à direita do núcleo nominal (LUCCHESI e DÁLIA, no prelo).

Diante do resultado da variável número de constituintes do SN e sua relação com a configuração sintagmática, nos parece ser necessário uma ampliação da investigação sobre a estrutura sintagmática do PB popular, tendo em vista que enquanto outros trabalhos apontam para um princípio de simplicidade, neste trabalho esse princípio não se confirma. Nesta tese, o resultado aponta para a semântica de quantificação do elemento constituinte do SN como contexto favorecedor à aplicação da CN.

### 5.2.3 Variável *número de constituintes flexionáveis do SN*

Além de analisar o número de constituintes do SN como um todo, analisei o número de constituintes flexionáveis do SN. Para essa variável, testei a hipótese de que quanto menor a quantidade de itens pluralizáveis, maior seria a probabilidade da aplicação da regra de CN. Do lado oposto, os SNs com número maior de constituintes flexionáveis teriam menos concordância.

Os resultados iniciais mostraram que havia 14 dados para o nível quatro ou mais constituintes, e que, em todos esses dados, não houve concordância, como no exemplo (23). Em função disso, amalgamamos os fatores dos SNs com três e quatro constituintes flexionáveis.

(23) nem com as ôtras meninas pequena [ROSF20]

Na Tabela 6, estão distribuídas as frequências e proporções dos dados para a variável número de constituintes flexionáveis do SN em relação à presença de CN no português popular do interior da Bahia. Podemos observar que os SNs com um item flexionável obtiveram percentual de 25,4% de aplicação da regra de concordância. Vale destacar que em todos os dados desse nível, o item não flexionável foi um numeral, o que reforça a influência da semântica de pluralidade de determinadas classes gramaticais, conforme o exemplo (24).

(24) há quinze anos atrás [ANTM25]

O percentual de aplicação da regra concordância para os SNs com dois constituintes flexionáveis foi de 14,8%, e 8,8% para os SNs com três ou mais constituintes flexionáveis.

Tabela 6. Frequências e proporções da variável *número de constituintes flexionáveis do SN* sobre a aplicação da CN.

<b>Número de constituintes flexionáveis do SN</b>	<b>Freq./Total</b>	<b>Proporção</b>
Um	340/1.340	25,4%

Dois	473/3.187	14,8%
Três ou mais	30/340	8,8%
<b>TOTAL</b>	<b>843/4.867</b>	<b>17,3%</b>

Fonte: elaborada pela autora.

O teste de qui-quadrado evidenciou que a diferença entre as proporções da variável número de constituintes flexionáveis do SN foi significativamente diferente de zero, com valor-p < 0,001, portanto essa variável foi incluída no modelo de regressão logística. O modelo de efeitos mistos, conforme Tabela 2, apontou correlação significativa dessa variável com a CN, pois em todos os seus níveis o valor-p foi menor que 0,05.

As estimativas de *logodds* indicaram que o favorecimento da aplicação da CN é maior em 0.738 nos SNs com um item flexionável, enquanto que para os SNs com dois, três ou mais itens flexionáveis as estimativas de aplicação da regra são menores, -1.013 e -0.903 *logodds*, nesta ordem. Através desse resultado, a hipótese levantada para essa variável é confirmada: quanto menor a quantidade de itens pluralizáveis, maior a probabilidade da aplicação da regra de CN.

Na análise sintagmática para o português afro-brasileiro (ANDRADE, 2006), essa variável não foi selecionada no modelo estatístico, e nos estudos de Baxter (2009) e Lopes (2001) essa variável não foi controlada, já que esses estudos apresentaram análises mórficas da CN. O resultado obtido nesta análise segue a proposição de que o desempenho linguístico tende a apresentar uma harmonia formal entre os seus elementos ou quando há a presença de um elemento com valor semântico de plural, como os numerais (SCHERRE, 1988), conforme exemplo (25). O que corrobora, mais uma vez, o resultado da variável configuração sintagmática do SN, cujas estruturas com mais concordância foram aquelas que continham algum elemento com ideia de quantidade.

- (25) a. um menino de nove dez anos [ANTM25]  
 b. agora que vai interá quatro anos [ANTM25]

Foram feitos testes estatísticos de cruzamentos de dados, porém o R não detectou interações entre as *variáveis configuração sintagmática do SN* e *número de constituintes flexionáveis do SN*. Pode ser que isso ocorresse devido à distribuição não

ortogonal da quantidade de dados entre os fatores das variáveis, logo, não foram viáveis interações confiáveis com quantificação.

#### 5.2.4 Variável *realização do núcleo*

A análise dessa variável girou em torno de verificar a influência da presença/ausência de um núcleo formado por nome no sintagma como condicionamento para a regra da CN. A hipótese é que a ausência do núcleo formado por nome favoreça a CN devido à função textual do SN. A estrutura sintagmática pode ter relação com ideias fora do texto de que maneira quando há algum tipo de relação espacial, tempo, medida e valor, o falante tende a marcar o plural para reforçar ao que se refere (SCHERRE, 1988).

Os resultados apresentados na Tabela 7 mostram que os SNs sem núcleo tiveram um percentual de 22,9% de aplicação da CN no português popular do interior da Bahia, enquanto os SNs com núcleo apresentaram um percentual de 17,3% de aplicação da regra. Infiro que isso tenha ocorrido porque, a depender do contexto da informação da sentença, o falante pode ter a necessidade de inserir as marcas de plural no SN. Além disso, o falante pode evitar a repetição do nome referente, se este fora mencionado anteriormente por ele ou na fala do documentador, como mostra o exemplo (26), e o SN pode ser constituído sem núcleo incorporando dados indefinidos nominalmente, como consta no exemplo (27).

(26) Os meus cansavam [ESLF57]

(27) eu também sei respeitá os mais velho [VALM46]

Tabela 7. Frequências e proporções da variável *realização do núcleo* sobre a aplicação da CN.

<b>Realização do núcleo</b>	<b>Freq./Total</b>	<b>Proporção</b>
SNs sem núcleo	8/35	22,9%
SNs com núcleo	835/4.832	17,3%
<b>TOTAL</b>	<b>843/4.867</b>	<b>17,3%</b>

Fonte: elaborada pela autora.



O resultado da análise univariada indicou um valor-p 0.5192, cuja interpretação leva a considerar que essa variável ficou no limite da significância em relação à variável dependente. Embora o valor-p do teste de qui-quadrado não tenha sido tão significativo, a variável foi incluída na regressão logística para verificar uma possível correlação multivariada. O modelo de efeitos mistos indicou que não houve correlação da realização do núcleo com a CN quando combinada com as demais variáveis em estudo.

Em um estudo recente, Lucchesi e Dália (2020) associaram a realização do núcleo com a posição do determinante no SN e, apesar do foco da análise desses autores ter sido o papel do determinante no sintagma, eles sugerem que a (não) marcação do plural quanto à CN pode estar associada ao papel do determinante, pois há contextos morfossintáticos em que o determinante pode se comportar como núcleo nos SNs de acordo com a posição que ocupam (cf. LUCCHESI E DÁLIA, 2020, p. 790-791). Como, na presente pesquisa, a variável realização do núcleo não foi estaticamente correlacionada com as demais, essa perspectiva de análise em conjunto com o determinante não foi considerada, abrindo espaço para trabalhos posteriores em que a quantificação dos dados permita mais análises estatísticas.

### **5.2.5 Variável *saliência fônica***

A variável *saliência fônica* segue a hipótese de que quanto mais saliente for a diferença entre singular e plural, mais provável seja a aplicação da regra de concordância (SCHERRE, 1988; ANDRADE, 2006), sendo que analisei SNs em que todos os itens sejam regulares e SNs em que, pelo menos um item, apresente diferenciação fônica quando pluralizado.

A hipótese para essa variável é de que a CN é favorecida nos SNs quando há algum item cuja *saliência fônica* seja mais evidente na oposição singular/plural. A Tabela 8 exibe os resultados encontrados em relação às frequências e proporções dos níveis dessa variável quanto ao emprego da concordância, que teve um percentual maior nos SNs que contém algum item com plural irregular, 34,7% de aplicação da regra, a exemplo de (28), em que a palavra *novo* possui plural duplo *novos*, logo, a *saliência morfofonológica* é mais perceptível.

(28) e agora já tem ôtos novos amigos [JOSM28]

Tabela 8. Frequências e proporções da variável *saliência fônica* sobre a aplicação da CN.

<b>Saliência fônica</b>	<b>Freq./Total</b>	<b>Proporção</b>
<i>SN com saliência fônica</i>	261/753	34,7%
<i>SN sem saliência fônica</i>	582/4.114	14,1%
<b>TOTAL</b>	<b>843/4.867</b>	<b>17,3%</b>

Fonte: elaborada pela autora.

Conforme o exemplo (29), a saliência fônica é mais perceptível em itens lexicais cujo plural é duplo, em que a proeminência da sílaba tônica é mais evidente na oposição singular/plural. Além desse nível, que é o mais explícito segundo a escala de saliência proposta por Scherre (1988), o emprego da CN ocorreu nos demais níveis da hierarquia. Em (a), vemos o item “reais”, cuja forma singular é terminada em -l e, ao passar para o plural, ocorre a supressão do -l com a inserção de -is; em (b), há o exemplo do item terminado em -ão no singular, cujo plural se faz com -es ou com a inserção de -s; o exemplo (c) contém um item com singular terminado em -r e com o plural terminado em -es; elencada em (d) está uma ocorrência referente ao nível pouco saliente, um item terminado em -s ou -z, cujo plural é feito com -es.

- (29) a. Só um quilo de carne tá custando sete **reais** [VALM46]  
 b. as **condições** num... num dá [MAGF26]  
     tem dois **irmãos** [MANM76]  
 c. Aí conheci muitos **lugares** [MAGF26]  
 d. muitas **vezes** aqui [MARF30]

Em todos os contextos de saliência fônica apresentados nos exemplos em (29) se verifica a concordância. E, para os mesmos contextos, houve também a ausência de concordância, ou seja, não há um contexto categórico que leve ao emprego da regra, como se vê nos exemplos em (30), os quais seguem a mesma ordem em relação à escala de saliência, do mais saliente para o menos saliente.

- (30) a. pra dá os **porco** [ARLM71]  
 b. os ôtos **animal** [ISRM52]  
 c. as **condição** de primêra era precária [ANTM25]  
     esquentô as **mão** [ANTM81]  
 d. aí fica difícil pros **morado** [JOJM28]  
 e. Já fui lá três **vez** [JOJM28]

Na análise univariada, a saliência fônica fez um resultado significativo, valor  $p < 0,001$ , o que levou a sua inclusão na análise multivariada. A regressão logística apontou que houve correlação da saliência fônica com outras variáveis.

Os *logodds* do modelo de efeitos mistos foram estimados em -1.013 para SNs com saliência e -1.076 para SNs sem saliência, ambos os valores em desfavorecimento quanto à aplicação da regra de CN. Esse resultado levou a uma observação mais detalhada dos dados e verifiquei que os itens lexicais salientes coletados na amostra pertencem, em sua grande maioria, ao nível 2 da escala de saliência, sendo a ocorrência mais frequente o item *reais* em estruturas de SN constituído por [numeral + nome], como os exemplos em (31).

- (31) a. pagano é cinco, seis *reais* a diára  
 b. num dé dois *reais* por dia  
 c. faço de quatro *reais*

Retomando o que fora ilustrado no Quadro 1 – níveis de saliência fônica estabelecidos por Scherre (1988), fiz um levantamento das ocorrências dos itens lexicais correspondentes a esses níveis no *corpus* deste trabalho. No Quadro 12, podemos ver, na ordem dos itens lexicais mais salientes para os menos salientes, que o nível 2 - itens terminados em *-l* com plural terminados em *-is* – foi o nível com a maior frequência de dados.

Quadro 12. Níveis de saliência fônica (SCHERRE, 1988) e as ocorrências no português popular do interior da Bahia.

Níveis de saliência fônica	Número de ocorrências/ exemplo
1. itens com plural duplo	16 ocorrências tomava conta dos porcos
2. itens terminados em <i>-l</i> com plural terminados em <i>-is</i>	123 ocorrências uns policiais veio trazê ela aqui
3. itens terminados em <i>-ão</i> , cujo plural se faz com inserção de <i>-s</i> e <i>-es</i>	36 ocorrências Dos seis irmãos faz as adorações lá
4. itens terminados em <i>-r</i> , cujo plural se faz através da inserção de <i>-es</i>	26 ocorrências Os vereadores, devia elegê

5. itens terminados em -s com plural terminados em -es	60 ocorrências levô oito meses
--	-----------------------------------

Fonte: elaborado pela autora.

Os níveis mais propícios à aplicação da CN são os níveis 1 e 2 porque envolvem mudanças diversas na forma plural: mudanças vocálicas mais perceptíveis aos falantes. Como ocorrência dos itens lexicais de nível 1 é muito baixa em relação aos demais, suponho que os falantes da amostra não façam diferenciação entre itens mais salientes e os menos salientes justamente por utilizarem itens lexicais com saliência menos explícita, principalmente porque não houve interação significativa entre as variáveis saliência fônica e escolaridade, conforme teste aplicado no modelo estatístico.

No que tange ao nível 2, que é considerado bem saliente, verifiquei uma ocorrência expressiva da palavra *reais*: das 123 ocorrências gerais do nível 2, 116 são com a presença do item *reais*<sup>34</sup> após um numeral. Essas ocorrências podem se tratar de expressões fixas, que são consideradas como “um grupo cristalizado com um determinado valor lexical preenchido convencionalmente, ou seja, ligado a um fato social” (XATARA e OLIVEIRA, 2002. p. 57) ou ainda como uma “sequência de palavras memorizadas pelos falantes como um todo unitário, sendo igualmente recuperada da memória em bloco, sem o intermédio obrigatório da aplicação de regras de valor geral” (FULGÊNCIO, 2008, p. 101).

Além disso, há o fato de a escolarização, como veremos adiante, não exercer um papel relevante na comunidade de fala em estudo, uma vez que falantes escolarizados tendem a perceber mais as diferenciações morfofonológicas do que falantes menos escolarizados.

Nos estudos revistos para esta pesquisa, a saliência fônica apresentou valores estatísticos que indicam o condicionamento bastante favorável à aplicação da marcação de plural e, conseqüentemente, aplicação da regra de CN nas formas que se apresentam mais perceptíveis ao falante na oposição singular/plural quando se tratou de variedades do português urbano (SCHERRE, 1998; LOPES, 2001; BRANDÃO e VIEIRA, 2012).

---

<sup>34</sup> Ressalto a importância de um estudo posterior com a análise linguística das ocorrências dos itens *real/reais* no *corpus* do português popular do interior da Bahia.

Esses resultados diversos permitem dizer que a saliência fônica, apesar de ser uma variável importante nos estudos sobre concordância, pode atuar de forma diferente de acordo com as características sociais dos falantes, das comunidades, como vimos neste trabalho para o português do interior da Bahia, que possui características sociais de variedade rurbana (ANTONINO, 2007).

### 5.2.6 Variáveis *classe gramatical do constituinte marcado e posição da marca de plural*

Devido à quantidade expressiva de dados de SNs sem concordância entre seus constituintes, considereei necessário inspecionar de forma descritiva a ausência nos 4.024 SNs na perspectiva mórfica, em que cada constituinte do SN é uma ocorrência. Para isso, foram codificados cada item do SN em que a marca formal de plural fora expressa, mesmo que no SN como um todo a concordância fosse ausente, conforme exemplificado em (32).

- (32) a. muitos filho obediente  
 b. os pai

A análise da variável *classe gramatical do constituinte marcado* busca verificar qual a classe que mais recebe marca de plural, sendo que no PB os determinantes costumam ser marcados quase categoricamente pelo fato de encontrarem-se em posição à esquerda do núcleo (GUY, 1981; LOPES, 2001; BRANDÃO, 2013; OUSHIRO, 2015). A Tabela 9 mostra os resultados.

Tabela 9. Resultado descritivo da *classe gramatical do constituinte* na primeira posição do SN.

<b>Classe gramatical do constituinte</b>	<b>Nº de ocorrências</b>	<b>Proporção<sup>35</sup></b>
Artigos	1.799	44,7%
Pronomes demonstrativos	417	10,3%
Quantificadores indefinidos	274	6,9%
Pronome possessivos	224	5,5%
Adjetivos	7	0,1%

Fonte: elaborada pela autora.

<sup>35</sup> A proporção indicada refere-se ao total geral dos SNs com ausência de concordância, 4.024 dados.

Conforme a Tabela 9, em que são elencadas as classes de palavras cujos itens estão em primeira posição do SN, os artigos definidos e indefinidos foram os itens que mais apresentaram a marca de plural -s, nos SNs sem concordância em sua totalidade, com um percentual de 44,7% do total dos dados. Os exemplos (33) e (34) evidenciam isso.

(33) os menino [ANTM65]

(34) umas festa boa [ANTM25]

A classe dos pronomes também se mostrou mais frequente e com os seguintes percentuais de marcação: pronomes demonstrativos com 10,3% e pronomes possessivos com 5,5%. Seguem os exemplos (35) e (36), respectivamente. Vale salientar que, a maioria dos itens pertencentes essas classes gramaticais, principalmente os artigos, possuem a função de determinantes, contexto bastante favorecedor para a inserção da marca de plural.

(35) esses prefeito nunca ajudaro em nada [VALM46]

(36) meus filho eu criei trabalhano muito [VALF55]

Os quantificadores indefinidos perfizeram 6,9% de marcação mórfica de plural, e os adjetivos apenas 0,1%, como mostram os exemplos (37) e (38).

(37) já vinha isso era de muitos tempo [JOSM28]

(38) um lugá pa tê boas festa [ANTM25]

Para os SNs em que os constituintes receberam a marca de plural em primeira e segunda posições, as frequências foram baixas, pois estruturas de SN com mais constituintes também são pouco frequentes no *corpus* mediante o total geral dos dados, como vimos os resultados da variável configuração sintagmática do SN. A Tabela 10, a seguir, apresenta as classes marcadas em SNs com mais de dois constituintes, mas com ausência de concordância na perspectiva sintagmática.

Tabela 10. Resultado descritivo da *classe gramatical do constituinte* na primeira e na segunda posição do SN.

<b>Classe gramatical do constituinte</b>	<b>Nº de ocorrências</b>	<b>Proporção<sup>36</sup></b>
Artigo + pronome possessivo	27	0,6%
Artigo + quantificador indefinido	17	0,4%
Quantificador indefinido + artigo	10	0,2%
Artigo + substantivo	9	0,2%
Artigo + adjetivo	6	0,1%
Pronome demonstrativo + substantivo	4	0,09%
Quantificador indefinido + substantivo	4	0,09%

Fonte: elaborada pela autora.

Os resultados mostram que as classes mais marcadas na primeira e na segunda posição do SN foram: artigo + pronome possessivo: 0,6%; artigo + quantificador indefinido: 0,4%; quantificador indefinido + artigo e artigo + substantivo, ambas com 0,2%. Para ilustrar os fatores da Tabela 10, seguem os exemplos de (39) a (45) na mesma ordem em que foram elencados na Tabela.

(30) moço, isso aí é coisa dos seus tempo [ANTM25]

(40) só ficô uns certos mei pra cá [ISRM52]

(41) correu todos os hospital [VALF55]

(42) as pessoas fraca assim [GUIF26]

(43) umas boas festa [VALM46]

(44) daquelas crianças desnutrida [MACF45]

(45) tem vários lugares bonito [LUCF23]

De maneira geral, a classe mais marcada foi a classe dos artigos, mesmo em SNs com dois constituintes marcados. Os SNs com presença de numerais não foram considerados pelo caráter semântico de pluralidade dessa classe e por serem itens que não apresentam inserção ou supressão de morfema de plural.

<sup>36</sup> A proporção indicada refere-se ao total geral dos SNs com ausência de concordância, 4.024 dados.

Os adjetivos e os substantivos foram as classes que menos receberam marca de plural, uma vez que os adjetivos geralmente aparecem em posições pós-nucleares e os substantivos ocupam, majoritariamente, a posição de núcleo do SN. Nos estudos de Scherre (1988), Lopes (2001) e Baxter (2009), os resultados são semelhantes ao desta análise.

Para a variável *posição do constituinte marcado*, analisei se os constituintes em primeira posição do sintagma tendem a receber mais marcas de plural do que os demais constituintes (SCHERRE, 1988; LOPES, 2001; BAXTER, 2009). Apresento na Tabela 11 os resultados, salientando que os numerais também não foram inclusos na análise dessa variável porque são itens que possuem a carga semântica de plural evidente.

Tabela 11. Resultado descritivo da *posição do constituinte marcado*.<sup>37</sup>

<b>Posição do constituinte marcado</b>	<b>Nº de ocorrências</b>	<b>Proporção<sup>38</sup></b>
primeira posição	2.716	67,4%
primeira posição e segunda posição	90	2,2%

Fonte: elaborada pela autora.

Os resultados confirmam a hipótese e apontam similaridade com todos os estudos revistos para esta pesquisa, pois a marca de plural apenas na primeira posição foi mais frequente e apresentou um percentual de 67,4% de presença da marca formal de plural, conforme exemplos em (46).

- (46) a. se as autoridade num dá jeito [ANTM25]  
 b. vai brinca com aquelas armadilha [VALM46]  
 c. muitos ano morano aqui [VALF70]  
 d. quem qué construí suas casa [ANTM25]  
 e. de novos tempo pra cá [JOSM28]

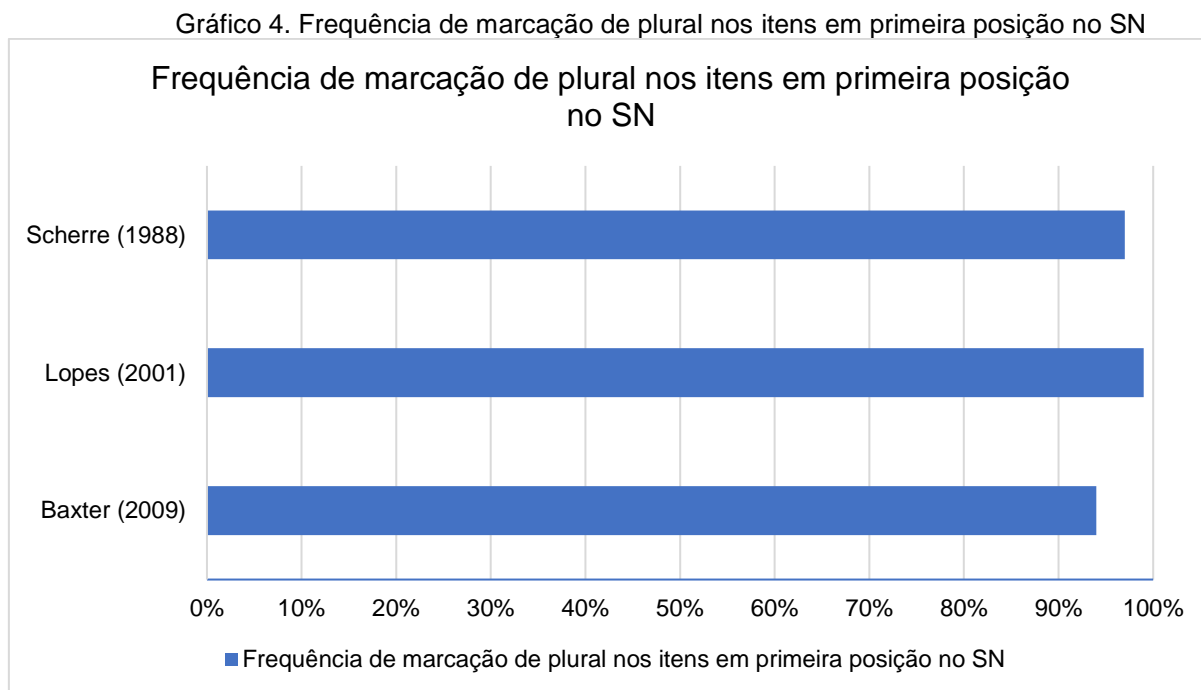
O Gráfico 4 exibe a semelhança dos resultados dos estudos para variedades urbanas do PB, Rio de Janeiro (SCHERRE, 1988), Salvador (LOPES, 2001), e para

<sup>37</sup> Inicialmente, estabeleci o nível marca de plural nas primeira, segunda e terceira posições. No entanto, dos 4.867 SNs sem concordância, apenas 1 dado desse nível foi encontrado. A saber, refere-se à ocorrência: “nem com as ôtras meninas pequena” [RORF20].

<sup>38</sup> A proporção indicada refere-se ao total geral dos SNs com ausência de concordância, 4.024 dados.



variedades rurais e afro do PB, Helvécia (BAXTER, 2009), Helvécia, Sapé, Rio de Contas e Cinzento (ANDRADE, 2003, 2006) com os resultados desta pesquisa.



Fonte: elaborado pela autora.

Segundo esses autores, no PB, os itens em primeira posição são mais frequentemente alvo da marca mórfica de plural porque inserem a informação nova de plural no sintagma, de maneira que, nas demais posições, a marca pode ser dispensada assim como ocorre neste trabalho com português do interior da Bahia.

Na seção 5.3, apresento os resultados da análise quantitativa e probabilística das variáveis sociais em relação à aplicação da regra de CN.

### 5.3 O ENCAIXAMENTO SOCIAL DA CN

Por entender que os fatores sociais estão interligados aos fatores linguísticos da CN, e pela fundamentação teórica assumida, considero as seguintes variáveis para a análise sintagmática desse fenômeno: *sexo*, *faixa etária*, *escolaridade*, *estadia fora da comunidade e localidade*. Nas subseções seguintes, apresento os resultados obtidos através da análise multivariada de regressão logística para cada uma das variáveis.

### 5.3.1 Variável sexo

Conforme discussão estabelecida no capítulo 4, subseção 4.3.3, os estudos sociolinguísticos seguem, geralmente, duas tendências: (i) quando se trata de comunidades de fala urbanas, as mulheres costumam utilizar mais a variante de prestígio do que os homens; (ii) já nas comunidades rurais, em sua maioria, os homens é que fazem mais uso da variante padrão (LUCCHESI, 2009). No entanto, o comportamento linguístico de homens e de mulheres nem sempre segue essas tendências, pois é necessário considerar o cenário sociocultural e econômico em que os falantes estão inseridos.

Sobre o encaixamento da mudança linguística no contexto social, Labov (2008 [1972]) diz que o papel das mulheres na mudança ocorre não somente pela sensibilidade feminina às formas de maior prestígio social, mas também pela diferenciação de papéis de homens e mulheres na sociedade, ligada aos padrões de interação social na vida diária. Os homens, por sua vez, podem liderar a mudança linguística por estarem mais inseridos no mercado de trabalho do que as mulheres, também em decorrência dos papéis de gênero.

Dados do IBGE (2018b), sobre as estatísticas de gênero no Brasil, apontam que, apesar de lento, o processo de mudança dos padrões culturais de gênero proporcionou a entrada das mulheres no mercado de trabalho, reduziu a taxa de fecundidade e elevou continuamente os níveis de escolaridade das mulheres a partir da década de 1980. Embora a estrutura ocupacional e financeira de homens e mulheres permaneça muito desigual, vários indicadores mostram uma tendência geral de maior nível de escolarização das mulheres em relação aos homens. Soma-se a isso o fato de que, em linhas gerais, há uma entrada precoce dos homens no mercado de trabalho e muitos deles não prosseguem nos estudos. Todavia, dada a complexidade da realidade brasileira, esses dados podem variar bastante de acordo com a região do país, principalmente no que diz respeito às diferenças entre os grandes centros urbanos e as cidades mais interioranas.

Neste estudo, levando em conta as configurações socioeconômicas sistematizadas sobre o interior do Estado da Bahia (ANTONINO, 2007; MENDES, 2016; SANTOS, 2016), a hipótese é de que os homens utilizem mais a variante padrão

do que as mulheres por terem mais ocupações fora da comunidade, em busca de emprego e melhores condições de vida, enquanto as mulheres tendem a permanecer nos serviços domésticos e na criação dos filhos.

A Tabela 12 expõe os resultados das frequências e proporções da influência da variável sexo sobre a aplicação da regra de CN. Os resultados sugerem uma leve, mas perceptível diferença no percentual de aplicação da CN entre homens e mulheres, com elas empregando mais a regra com 20,5% de aplicação, enquanto os homens aplicaram a CN em cerca de 14, 1% dos dados.

Tabela 12. Frequências e proporções da variável sexo sobre a aplicação da CN.

<b>Sexo</b>	<b>Freq./Total</b>	<b>Proporção</b>
<i>Masculino</i>	337/2.397	14,1%
<i>Feminino</i>	506 /2.470	20,5%
<b>TOTAL</b>	<b>843/4.867</b>	<b>17,3%</b>

Fonte: elaborada pela autora.

O teste de qui-quadrado, aplicado para uma análise univariada, indicou um valor-p de 3.975e-09, apontando diferença significativa entre as proporções da variável sexo. A regressão logística indicou que houve correlação da variável sexo com a variação da CN. Na Tabela 2, o modelo de efeitos mistos mostra a estimativa de -0.651 *logodds* para os homens em relação ao nível do intercept, -1.013 *logodds*, que se refere às mulheres. Embora ambos os sexos apresentem pouco emprego da regra, o sexo masculino desfavorece menos a aplicação da CN do que o sexo feminino mulheres no português do interior da Bahia.

Esse panorama é oposto ao evidenciado no estudo de Andrade (2006) para as comunidades afro-brasileiras baianas, nas quais as mulheres aplicam mais a concordância do que os homens. Também é oposto aos estudos de Scherre (1988), no Rio de Janeiro; e Lopes (2001) em Salvador, ambas comunidades urbanas. Segundo as autoras, nessas comunidades, as mulheres possuem uma maior sensibilidade ao uso de regras socialmente prestigiadas devido às pressões sociais, como ascensão e permanência no mercado de trabalho, o que não se configura para a comunidade de fala desta pesquisa.

### 5.3.2 Variável *faixa etária*

A análise da variável faixa etária revela o aspecto geracional no condicionamento do fenômeno da CN, conforme evidenciado na subseção 4.3.5 do Capítulo 4. A expectativa para essa variável foi de que os mais jovens empregassem mais a CN do que os informantes das outras faixas etárias em função do maior acesso à escola e da maior possibilidade de contato com a variante de maior prestígio social. Os resultados são expostos na Tabela 13 a seguir, mostram que o percentual de emprego da CN é maior entre os informantes jovens, 28%, do que entre os informantes da faixa etária intermediária e idosa, 14, 1% e 11,5%, nessa ordem.

Tabela 13. Frequências e proporções da variável *faixa etária* sobre a aplicação da CN.

<b>Faixa etária</b>	<b>Freq./Total</b>	<b>Proporção</b>
Jovem (25 a 35 anos)	406/1.451	28%
Intermediária (45 a 55 anos)	242/1.714	14,1%
Idoso (mais de 65 anos)	195/1.702	11,5%
<b>TOTAL</b>	<b>843/4.867</b>	<b>17,3%</b>

Fonte: elaborada pela autora.

A análise univariada do teste de qui-quadrado indicou um valor-p significativo, 2.2e-16, para a variável faixa etária. Ao inclui-la na regressão logística, o modelo de efeitos mistos apontou que houve uma certa correlação da faixa etária com a CN em combinação com as demais variáveis em análise. As estimativas em *logodds* – mostradas na Tabela 2 – indicaram que a faixa etária jovem favorece a aplicação da regra de CN, 1.013 *logodds*, enquanto as demais faixas etárias desfavorecem a aplicação da regra: -0.009 *logodds* para faixa intermediária e -1.013 para a faixa idosa.

Os resultados confirmam a hipótese de que os falantes mais jovens realizam mais concordância do que os mais velhos, além de indicar que a concordância tem um padrão aquisicional. Esse comportamento é diverso nas diferentes variedades do PB, por exemplo: no Rio de Janeiro (Scherre, 1988) e em Salvador (Lopes, 2001) há evidência de variação estável; nas comunidades afro-brasileiras baianas (Andrade,

2003, 2006; Baxter, 2009), a variável faixa etária sugere processo de mudança aquisicional, semelhantemente ao panorama desta pesquisa.

### 5.3.3 Variável escolaridade

Conforme já mencionado, o papel da escolaridade no tocante à aquisição da norma padrão é bastante relevante, principalmente quando estão em variação linguística uma variante de prestígio social e uma variante estigmatizada, como é o caso do que ocorre com o fenômeno da CN, em que a ausência de concordância se encontra no contexto de estigma social (LUCCHESI, 2000; 2015; LUCCHESI e DÁLIA, 2021). Em virtude disso, a hipótese para a variável escolaridade foi de que os falantes com algum nível de escolaridade apresentassem mais emprego da regra de CN do que aqueles que nunca frequentaram a escola.

A Tabela 14 expõe que a proporção de aplicação da CN no português do interior da Bahia foi de 22, 2% para os falantes com algum nível de escolaridade, ainda que mínimo, e de 10,4% para os falantes que nunca frequentaram a escola. O valor-p indicado pelo teste de qui-quadrado,  $2.2e-16$ , revelou que a diferença entre as proporções da variável foi significativa.

Tabela 14. Frequências e proporções da variável *escolaridade* sobre a aplicação da CN.

<b>Escolaridade</b>	<b>Freq./Total</b>	<b>Proporção</b>
Pouca escolaridade	636/2.868	22,2%
Nenhuma escolaridade	207/1.999	10,4%
<b>TOTAL</b>	<b>843/4.867</b>	<b>17,3%</b>

Fonte: elaborada pela autora.

A inserção da escolaridade no modelo de efeitos mistos mostrou que, quando combinada com os fatores de outras variáveis, não houve correlação com o fenômeno da CN. Para avaliar mais especificamente a interferência da escola no comportamento da CN no português do interior da Bahia, realizei testes de interação com as outras variáveis sociais na regressão logística, que indicou que não houve nenhuma interação significativa. Esse resultado sugere que a escola não exerceu um papel tão

considerável quanto à variante padrão da CN na comunidade de fala em estudo. Isso se verifica através das informações sobre o *corpus*, as quais mostram que os informantes que frequentaram a escola não chegaram a completar o ensino fundamental, e que, provavelmente, essa frequência não foi o suficiente para determinar uma avaliação acerca das variantes mais aceitas ou não socialmente.

Andrade (2006), que analisou o português rural afro-brasileiro, diz que essa variável apresenta conformidade com o papel social da escola nas comunidades rurais afro-brasileiras:

Embora os resultados fujam ao padrão, com pessoas sem escolaridade alguma realizando mais a regra do que as que freqüentaram a escola: 10% de freqüência contra 7% do fator semi-analfabeto, consideramos que eles refletem a realidade, na medida em que, como ressaltamos, a escola não exerce, nas comunidades, influência no que diz respeito ao desenvolvimento dialetal dos adultos. Em primeiro lugar, muitos dos informantes da amostra que foram classificados como semi-analfabetos não tiveram um ano sequer de estudo. Podemos levar em conta, também, que até mesmo as instituições públicas de ensino da zona urbana funcionam precariamente e com um quadro de profissionais, muitas vezes, mal formados ou até mesmo leigos que não dominam de forma efetiva a variedade culta da língua portuguesa, não podendo, dessa forma, ampliar o universo dialetal de seus alunos, e, na zona rural esse quadro é ainda mais grave. Mesmo que se encontrem, em zonas afastadas, profissionais capacitados, não podemos obscurecer o fato de que muitos alunos não conseguem aprender, simplesmente porque não entendem a variedade lingüística de seu professor (ANDRADE, 2006, p. 208).

Para investigar de maneira mais específica a interferência da escola na (não)aplicação da CN no PB popular do interior da Bahia ainda testei um cruzamento de frequências e proporções das variáveis *escolaridade* e *faixa etária*, e foi possível verificar uma informação relevante: os jovens possuem mais escolaridade que os falantes da faixa intermediária e dos idosos. A leitura da Tabela 15 mostra que, embora não tão relevante na análise multivariada, a escolaridade entre os jovens reforça a inferência que a instrução formal, ainda que mínima, possui um papel relevante nos estudos sobre variação linguística, uma vez que os jovens foram os que mais frequentaram a escola e os que mais aplicaram a CN.

Tabela 15. Cruzamento da *faixa etária* com *escolaridade*

FAIXA ETÁRIA	25-35 anos		45-55 anos		+ 65 anos		TOTAL		TOTAL GERAL Nº de ocor.
	Pres. Nº %	Aus. Nº %	Pres. Nº %	Aus. Nº %	Pres. Nº %	Aus. Nº %	Pres. Nº %	Aus. Nº %	
<b>ESCOLARIDADE</b>									
<b>POUCA ESCOLARIDADE</b>	393 14%	959 33%	177 6%	860 30%	66 2%	413 14%	636 22%	2.232 78%	2.868
<b>NENHUMA ESCOLARIDADE</b>	13 0,3%	86 4%	65 3%	612 30%	129 6%	1.094 55%	207 10%	1.792 90%	1.999
<b>TOTAL</b>	406 48%	1.045 26%	242 29%	1.472 36%	195 23%	1.507 37%	843 17%	4.024 83%	4.867

Fonte: elaborada pela autora.

### 5.3.4 Variável *estadia fora da comunidade*

A finalidade dessa variável foi verificar a confirmação ou não da hipótese de que os falantes que viveram certo período de tempo fora da comunidade fariam mais concordância do que aqueles que sempre viveram na comunidade (LUCCHESI, 2009). A Tabela 16 mostra que os informantes que viveram fora da comunidade por, pelo menos, 6 meses, aplicaram a CN em 18,7% das ocorrências, enquanto os que nunca viveram fora perfizeram 16,2% de emprego da concordância.

Tabela 16. Frequências e proporções da variável *estadia fora da comunidade* sobre a aplicação da CN.

<b>Estadia fora da comunidade</b>	<b>Freq./Total</b>	<b>Proporção</b>
Nunca viveu fora da comunidade por mais de 6 meses	428/2.645	16,2%
Viveu pelo menos 6 meses fora	415/2.222	18,7%
<b>TOTAL</b>	<b>843/4.867</b>	<b>17, 3%</b>

Fonte: elaborada pela autora.

A análise univariada com o teste de qui-quadrado indicou um valor -p significativo (0.02423), que levou à inclusão da estadia fora da comunidade na

regressão logística. No entanto, no modelo de efeitos mistos da regressão logística, com a ação simultânea de todos os demais fatores relevantes, essa variável não foi considerada como estaticamente relevante.

As proporções mostradas na tabela 16 indicam que os falantes que viveram pelo menos 6 meses fora da comunidade aplicaram mais a norma de prestígio. A inclinação ao emprego da regra de CN por esses falantes corrobora a influência externa na mudança linguística (Baxter, 2009). Nas entrevistas do *corpus* deste trabalho, há relatos de informantes que viveram em grandes centros urbanos, como Salvador e São Paulo, para trabalhar e, posteriormente, retornaram para suas comunidades, exemplo (47). Além disso, tanto os informantes que nunca viveram fora da comunidade quanto os que viveram, relataram que possuem contato frequente com familiares que moram nesses centros urbanos, exemplo (48).

(47) DOC: A senhora já morô algum tempo, assim, fora d'aqui de Poções? Algum lugá aqui na região ou [alguma] outra cidade? Passô a vim pr'aqui de fora, assim, daqui ou sempre morô aqui?

INF: Não, sempre eu morei aqui. Teve... só teve uma época que eu... que eu fiquei, morei um ano ali, perto de Jequié.

DOC: Um ano, né?

INF: É.

DOC: E que... por que motivo, assim, a senhora morô lá?

INF: Pra trabalhá.

DOC: Trabalho, né?

INF: É. [JUGF51]

(48) INF: Aqui o... aqui era uma cidade pequena, mas depois foi aumentano, aumentano, e lá vai. Agora eu gosto muito daqui, falá a verdade, eu num gosto de... Olha, eu vô em Salvador, na casa de minha filha, na casa de minhas irmã... [MASM76]

Esses dados sugerem que o deslocamento para áreas mais urbanizadas pode fazer com que os falantes incorporem à sua fala características mais próximas ao PB culto.

### 5.3.5 Variável *localidade*

Os municípios de Poções e Santo Antônio de Jesus são formados por zona rural e zona urbana, cujas principais diferenças se encontram no âmbito das atividades socioeconômicas desenvolvidas e nas redes de relações estabelecidas.



A análise da variável *localidade* teve como objetivo verificar se a hipótese de que os falantes da zona urbana aplicam mais CN do que os falantes da zona rural se confirma ou não, tendo em vista que a proximidade com grandes centros urbanos, como a capital do estado, Salvador, pode interferir no uso das variantes linguísticas de prestígio. Além disso, considero as relações rurbanas mensuráveis entre as características da rede social do falante e o seu comportamento linguístico.

Segundo Bortobi-Ricardo (2005), as redes de relações sociais envolvem a densidade e a multiplexidade. A densidade diz respeito à quantidade de pessoas que se conhecem entre si num determinado grupo de modo que, quando há um número grande de relações interpessoais dentro desse grupo, há uma densidade alta; por outro lado, quando há um baixo número de relações interpessoais, a densidade é caracterizada como baixa.

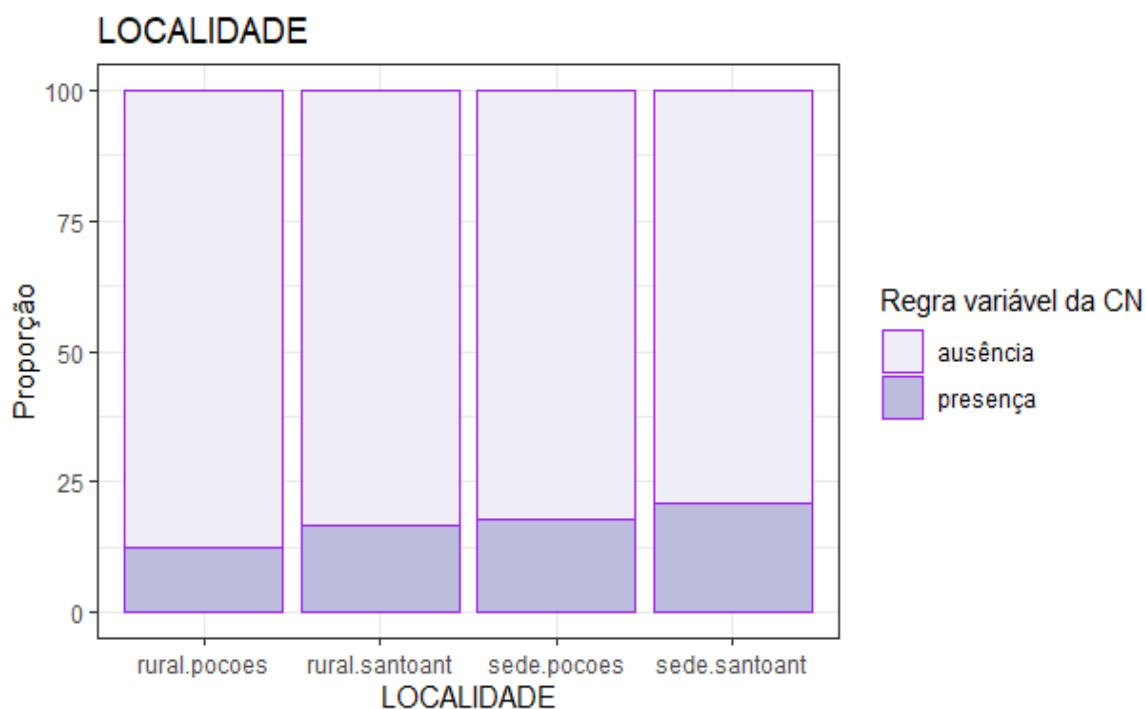
A multiplexidade pode se caracterizar de duas formas: (i) como uma única direção, sendo a rede de relações uniplex na qual há uma atuação de papéis mais restrita entre os indivíduos que a compõe, por exemplo, raramente se estabelece algum tipo de vínculo com os demais membros do grupo (médico/paciente; patrão/empregado); (ii) com direções múltipla, que é a rede social multipléxica, em que os indivíduos estabelecem interações diversas e simultâneas, como parentes, amigos, vizinhos etc. Bortoni-Ricardo (2005, p.84) afirma que “a alta densidade e a multiplexidade [...] são geralmente características dos sistemas tradicionais, isolados. Sistemas urbanos, por outro lado, tendem a uma densidade baixa e à ‘uniplexidade’”.

Seguindo esses pressupostos, a variável localidade para a análise da CN no interior da Bahia sugestiona que os falantes da zona rural dos municípios utilizem menos a regra padrão por estabelecerem predominantemente relações sociais de alta densidade e multiplex e estarem menos expostos à influência dos grandes centros urbanos. Já os falantes da zona urbana estariam mais expostos ao emprego da regra padrão por viverem nas sedes dos municípios que são regiões mais desenvolvidas, e por estabelecerem rede de relações de baixa densidade e uniplex.

Inicialmente, rodei os dados separando os falantes de acordo com o município onde residiam: zona rural e zona urbana de Poções, zona rural e zona urbana de Santo Antônio de Jesus. As proporções indicaram uma sutil, mas sugestiva, diferença de percentual de aplicação de CN entre os informantes, a saber: zona rural de Poções: 12, 1%; zona urbana de Poções: 17,9%; zona rural de Santo Antônio de Jesus: 16,4%;

zona urbana de Santo Antônio de Jesus: 20,9%. O Gráfico 5 foi plotado para uma visualização desse resultado.

Gráfico 5. Proporções da variável *localidade* em relação à aplicação da regra de CN.



Fonte: elaborado pela autora.

A apreciação desse resultado nos levou a investigar quais possíveis processos socio-históricos poderiam ter influenciado a variação na concordância nos municípios estudados quanto às atividades rurais e urbanas.

Santo Antonio de Jesus, segundo dados do IBGE (2018a), passou por uma urbanização crescente nas últimas duas décadas no setor industrial de forma que a estrutura setorial está distribuída da seguinte forma: 5,62% para agropecuária, 21,30% para indústria e 73,08% para serviços e comércio. O comércio no centro da cidade, especificamente a feira livre, atrai milhares de pessoas de regiões circunvizinhas, pois é considerada a feira livre mais barata do Estado da Bahia. Ademais, a cidade recebeu, no mesmo período, polos educacionais universitários,

como a Universidade Federal do Recôncavo da Bahia (UFRB) e a Universidade do Estado da Bahia (UNEB).

Em Poções, destaca-se a agropecuária como principal atividade socioeconômica. O comércio, na zona urbana, é pequeno. Apesar de possuir um maior dinamismo econômico do que os municípios circunvizinhos, Poções não possui grande destaque nos aspectos da vida urbana, como grande fluxo de pessoas, mercadorias, capitais e informações. A sede do município possui indústrias de pequeno porte que englobam a produção de palmito, de doce de banana, de esmalte e de palmilhas (ALMEIDA, 2013). O fato de ser localizado próximo a dois municípios de médio porte e que exercem forte influência na região – Jequié e Vitória da Conquista – faz com que os moradores de Poções se desloquem para estes com a finalidade de buscar melhores serviços de saúde e educação.

Segundo Almeida (2013), existe a Associação de Estudantes de Ensino Superior de Poções – AEESP, uma parceria entre o poder público municipal e os estudantes, que oferece serviço de transporte diário para Vitória da Conquista para garantir o acesso desses estudantes ao ensino superior, além dos estudantes que fazem cursos pré-vestibulares e ainda, trabalhadores que se dirigem diariamente para Vitória da Conquista.

Embora não sejam drásticas, podemos perceber diferenças entre os modos de vida dos municípios de Poções e de Santo Antônio de Jesus, o que pode explicar o comportamento linguístico levemente distinto quanto à regra de concordância. Provavelmente, os falantes de Santo Antônio de Jesus empregam mais a concordância que os falantes de Poções devido à maior amplitude de redes sociais disponíveis.

Ainda há um grande destaque para o comércio e para o setor de serviços, impulsionados pela migração da população rural para a sede do município. O município de Santo Antônio de Jesus também é um reconhecido produtor de móveis que atrai diariamente a população das regiões vizinhas, inclusive de Salvador e região metropolitana.

Como um dos objetivos desta pesquisa é verificar se a CN se encontra num *continuum* linguístico entre as comunidades mais isoladas até a capital do Estado da Bahia, e o interior abrange, estrategicamente, os dois municípios, amalgamei os dados para a análise da variável localidade. Os resultados são apresentados na

Tabela 17 e mostram que os falantes da zona urbana empregaram a CN em 20, 1% dos dados, enquanto os falantes da zona rural empregaram em 14,4% dos dados.

Tabela 17. Frequências e proporções da variável *localidade* sobre a aplicação da CN.

<b>Localidade</b>	<b>Freq./Total</b>	<b>Proporção</b>
Zona rural	345/2.389	14,4%
Zona urbana	498/2.478	20,1%
<b>TOTAL</b>	<b>843/4.867</b>	<b>17,3%</b>

Fonte: elaborada pela autora.

A análise univariada indicou que as proporções da variável localidade foram significativamente diferentes entre si. Porém, ao ser incluída no modelo de efeitos mistos para análise multivariada, a variável localidade não foi selecionada como estaticamente relevante.

Testei o cruzamento de dados da variável localidade com outras variáveis sociais, como *faixa etária* e *escolaridade*. O programa estatístico não encontrou nenhuma interação entre essas variáveis, portanto, acredito que esse resultado segue o padrão já mencionado em outros estudos linguísticos: quanto mais próximo às áreas urbanas, maior é a tendência de aplicação das regras da gramática normativa da língua (LOPES, 2001; LUCCHESI, 2015), haja vista as proporções indicadas na tabela 17.

No entanto, há que se salientar a complexidade das relações rurais e urbanas, bem como da formação das cidades no Brasil como um todo. Urge a necessidade de ampliação de estudos socioantropológicos e linguísticos a fim de aprofundar o (não) uso de variantes da língua.

#### 5.4 SÍNTESE

Neste capítulo, apresentei os resultados quantitativos e qualitativos, bem como a análise descritiva dos dados de CN no português popular do interior da Bahia. A análise descritiva considerou a presença de concordância, enquanto fenômeno

variável, na comunidade de fala em estudo, referente aos municípios de Poções e Santo Antônio de Jesus.

Os resultados gerais da variável dependente revelaram um percentual de 17,3% de emprego de CN e 83,7% de ausência. Essa significativa diferença no percentual para a variável dependente apontou lacunas na destruição dos dados em relação às variáveis dependentes, ou seja, houve certa falta de ortogonalidade nos fatores inicialmente levantados para análise e, dessa forma, foi necessário tratar dos dados em relação às variáveis para que as rodadas e os testes estatísticos se fizessem confiáveis.

Após, o amalgamento dos níveis de algumas variáveis, foram realizados testes estatísticos para a análise univariada (qui-quadrado) e para a análise multivariada (regressão logística com efeitos mistos). Na análise univariada, apenas a variável *número de constituintes do SN* não se mostrou relevante, pois as proporções de seus níveis não foram significativamente de zero. Na análise multivariada, foram feitos dois modelos de regressão logística, tendo em vista a falta de ortogonalidade dos fatores das variáveis *configuração sintagmática do SN* e *número de constituintes flexionáveis do SN*.

Os resultados dos dois modelos de regressão logística (cf. Tabela 1 e Tabela 2) indicaram que as variáveis mais relevantes no processamento estatístico do fenômeno variável CN no português popular da Bahia foram: configuração sintagmática do SN, número de constituintes flexionáveis do SN, saliência fônica, sexo, faixa etária e escolaridade.

Apresentei as frequências e proporções de duas variáveis linguísticas para a descrição mórfica dos dados de ausência de CN para verificar qual a classe morfológica recebeu mais marca de plural e qual a posição ocupada pelo constituinte marcado. Os resultados apontaram que as classes morfológicas que funcionam como determinantes (artigos e pronomes) foram as mais marcadas, corroborando com o resultado de que os constituintes em primeira posição são mais marcados.

Saliento que a análise quantitativa evidenciou um comportamento aquisicional da regra de CN que parece se difundir no SN em que o nome é precedido de um numeral ou quantificador, enquanto o contexto mais resistente à marcação é aquele em que um determinante precede o nome. A variável saliência fônica apresentou um resultado distinto da maioria dos estudos sobre CN: os itens mais salientes

desfavoreceram a aplicação da regra. Resultado que pode estar relacionado com a presença de expressões cristalizadas.

Os resultados da análise deste trabalho apresentaram certa semelhança com o que já fora observado em comunidades afrodescendentes do interior da Bahia (Andrade, 2003, 2006; Baxter, 2009), cujo padrão encontrado foi de aquisição da norma de prestígio. Porém, nas variedades do PB do Rio de Janeiro (Scherre, 1988) e de Salvador (Lopes, 2001), o *status* da CN se mostrou em variação estável.

## 6. O PORTUGUÊS POPULAR DO INTERIOR DA BAHIA E AS LÍNGUAS BANTU

Esta tese, além de descrever os condicionamentos que levam a aplicação da regra de concordância, apontando para um quadro de variação estável ou mudança em curso em municípios do interior da Bahia, visa também avaliar os resultados procurando responder as seguintes questões:

1) A variação na aplicação da regra de concordância no PB poder estar relacionada ao contato linguístico ocorrido no período de colonização do Brasil?

2) Teriam as línguas de substrato, principalmente as da família bantu, que foram as línguas mais faladas em território baiano (LUCCHESI, 2002; GUY, 2005; PESSOA DE CASTRO, 2006; PETTER, 2006), interferido na marcação de plural apenas no primeiro elemento do SN, um fenômeno observado no PB?

Para responder a essas duas questões, apresento o comportamento linguístico de algumas línguas do grupo bantu e algumas línguas indígenas brasileiras em relação à flexão de plural e à CN, nas seções 6.1 e 6.2. Levando em consideração a predisposição do PB marcar o plural no primeiro elemento SN, na seção 6.3, testo a hipótese de Guy (2005) sobre a interferência das línguas bantu na tendência de no PB ocorrer marcação de plural, prioritariamente, na primeira posição do SN.

Nesse sentido, proponho que essa característica tenha se fixado após um processo de competição e seleção de traços entre o português e as línguas de substrato, ou seja, as línguas bantu, que foram amplamente faladas na Bahia (LUCCHESI, 2002; PESSOA DE CASTRO, 2006) e conseqüentemente, suas propriedades poderiam ser concorrentes fortes para a aquisição dessa característica do PB.

### 6.1 A CONCORDÂNCIA NOMINAL NAS LÍNGUAS BANTU

Nesta seção, descrevo as principais características da marcação de plural em algumas línguas bantu, bem como a CN é expressa, nessas línguas, com a finalidade de comparar o comportamento morfossintático desse fenômeno com os dados do português popular do interior da Bahia, analisados anteriormente no Capítulo 5 deste trabalho.

Segundo Pessoa de Castro (2006), a principal característica das línguas bantu

é o sistema de classes que funciona por meio de prefixos que se ordenam em pares (cl. 1/2, cl. 3/4, etc), para exprimir a oposição singular e plural dos nomes, o aumentativo, o diminutivo, o locativo, o infinitivo dos verbos, permitindo ainda delimitar o sentido desse mesmo nome, como no caso da cl. 1/2, com prefixos mu- / ba-, referentes a seres humanos, a exemplo de ba.ntu, plural de mu.ntu, homem, ou, então, muleke e mukama. Da classe ku, a dos termos verbais (ku- é semelhante ao todo infinitivo verbal do inglês, to speak, falar), temos, entre outros, ku.xila, dormir, ku.xinga, insultar, ku.babata, tatear, enquanto da classe ka-, a dos diminutivos, temos kalunga, kamundongo, rato pequeno (PESSOA DE CASTRO, 2006, p. 3)

É a marcação morfológica de prefixos nominais de classe através dos processos de adição, substituição ou subtração de morfemas que a CN é expressa nas línguas bantu (NGUNGA, 2004). Em (1) está exemplificado o processo morfológico de adição nas línguas Changana e Nyungwe, no qual a palavra possui um morfema de singular, e a esse é acrescentado o morfema de plural.

- (1) a. Changana: **yin** - dlu  
**SING.-NOME**  
 'casa'
- ti-** yin - dlu  
**PL-SING.-NOME**  
 'casas'
- b. Nyungwe: **ny**-anga  
**SING.-NOME**  
 'chifre'
- zi-ny**-anga  
**PL-SING.-NOME**  
 'chifres'

(Adaptado de NGUNGA, 2004, p. 105)

O processo de substituição é caracterizado pela mudança de sentido das palavras quando certos morfemas substituem outros, como se vê nos exemplos de (2) a (4), nas línguas Nyanga, Makonde e Ronga.

- (2) Nyanga: **m**-tengo  
**SING.-NOME**  
 'árvore'
- mi**-tengo



**PL.-NOME**  
'árvores'

- (3) Makonde: **mw-edi**  
**SING.-NOME**  
'mês'

**my-edi**  
**PL.-NOME**  
'meses'

- (4) Ronga: **dri-tluka**  
**SING.-NOME**  
'folha'

**ma-tluka**  
**PL.-NOME**  
'folhas'

(Adaptado de NGUNGA, 2004, p. 106)

Diferentemente do processo de adição, no qual o morfema de plural é acrescentado ao morfema de singular, conforme exemplos (2), (3) e (4), o morfema indicador de plural substitui a marca de singular e atribui à palavra o sentido de mais de um. Esse é o processo de substituição. Por fim, há o processo de subtração, em que o plural é formado através da eliminação da marca do singular e sem adição de nenhum outro prefixo. Os exemplos em (5), na língua Yao, evidenciam esse processo.

- (5) a. **lu-sulo** 'rio'  
sulo 'rios'
- b. **lu-saau** 'cavaco'  
saau 'cavacos'

(NGUNGA, 2004, p. 107)

Os prefixos, nas línguas bantu, geralmente são variáveis de acordo com a classe. Fazendo referência a Bleek (1862, 1869), Ngunga (2004, p. 108) diz que o conjunto de nomes com o mesmo prefixo ou com o mesmo padrão de concordância chama-se classe nominal. Segundo Ngunga, foi Bleek (1862, 1869) quem notou pela primeira vez que os nomes destas línguas se organizam de forma sistemática, em grupos, de acordo com seus prefixos ou tipo de concordância. Dando continuidade ao trabalho de Bleek, Meinhof (1910) apresentou uma lista de prefixos que pertenciam a vinte e uma classes do que, segundo ele, poderiam ser do proto-bantu. Apresento no

Quadro 13 os prefixos nominais, suas classes e seus significados, conforme Meinhof (1910).

Quadro 13. Prefixos nominais do proto-bantu, segundo Meinhof (1910).

<b>Classe</b>	<b>Prefixo</b>	<b>Significado</b>
1	mu-	singular de 2
2	va-	plural de 1
3	mu-	singular de 4
4	mi-	plural de 3
5	li-	singular de 6
6	ma-	plural de 5 e de 14 de substâncias ou coisas incontáveis
7	ki-	singular de 8
8	vî-	plural de 7
9	Ni	singular de 10
10	lî-ni	plural de 9 e de 11
11	lu-	singular de 10
12	tu-	plural de 11, de 13 e de 19
13	ka-	singular de 12 e, às vezes de 14
14	vu-	singular de abstratos, massa, coisas incontáveis, e singular de 4, de 6 e de 13
15	ku-	Infinitivo
16	pa-	locativo (situacional)
17	ku-	locativo (direcional)
18	mu-	locativo (de interioridade)
19	pî-	Diminutivos
20	yu-	depreciativos e diminutivos
21	ya-	aumentativos e depreciativos

Fonte: Adaptado de Ngunga (2004, p. 109)

Estudos posteriores ao de Meinhof (1910) - Werner (1919), Guthrie (1967) - verificaram a existência de outras classes nominais nas línguas bantu, bem como semelhanças e diferenças entre essas classes. Postularam, então, um grupo de seis classes e prefixos que parecem ocorrer em todas as línguas bantu, a saber:

- (i) **mu-**;
- (ii) **ba-**;
- (iii) **mw-**;
- (iv) **mi-**;
- (v) **î-**;
- (vi) **ma-**.

Ngunga (2004) afirma que o estudo sobre a semântica das classes nominais das línguas bantu é um tema frequente nas pesquisas devido à tentativa de estabelecer ainda mais relações entre a forma fonética dos prefixos e os significados dos nomes. Ele menciona que há predominância de nomes que podem se agrupar em categorias gerais e propõe um quadro dos prefixos nominais que são afixados a temas nominais, designando seus significados. O Quadro 14 expõe esses dados.

Quadro 14. Classes, prefixos nominais e significados nas línguas bantu.

<b>Classes</b>	<b>Prefixos</b>	<b>Significados</b>
1 e 2	mu- e ba-	seres humanos, principalmente.
3 e 4	mu e mi-	plantas, predominantemente.
5 e 6	i- e ma-	animais e frutas, sobretudo.
7 e 8	ki- e ba-	coisas, basicamente.
9 e 10	N- e N-	alguns seres do reino animal, e outros.
11 e 10	du- e N-	coisas longas, principalmente.
14	bu-	substâncias e abstratos.
15	ku-	nomes verbais; infinitivo verbal.

Fonte: Adaptado de Ngunga (2004, p. 114)

Quanto à função sintática, os prefixos podem controlar a concordância do nome e de todas as palavras que se afixam a ele sintaticamente (cf. NGUNGA, 2004, p. 119). No SN, o nome é o centro do sistema de concordância, o que Ngunga (2004) afirma ser uma característica central das línguas bantu, e a estrutura do SN é, predominantemente, nome + modificador: o nome é o núcleo e o modificador, quando ocorre, pode ser um adjetivo, um outro nome, um pronome possessivo, uma oração



Outra forma de CN nas línguas bantu relacionada à quantificação ocorre com a presença de numerais no SN. De acordo com Mondlane (2012), a concordância entre o nome e o numeral é estabelecida pelo prefixo da classe do nome quantificado, como podemos ver em (13) na língua Changana, em que ocorrem os prefixos *mu-* e *va-* das classes 1 e 2, respectivamente.

- (13) a. **mufundhisi mun´we**  
           **mu-**      fundhisi  **mu-**      n´we  
           **SING. CL. 1** NOME  **SING. CL. 1** NUM.  
                   ‘um pastor’
- b. **vafundhisi vanharhu**  
           **va-**      fundhisi  **va-**      nharhu  
           **PL. CL. 2** NOME  **PL. CL. 2** NUM.  
                   ‘três pastores’

(Adaptado de MONDLANE, 2012, p.12)

Mais uma característica dos numerais, nas línguas bantu, é que alguns deles podem ser classificados com a função de nome ou de adjetivo e, em algumas línguas podem vir antes ou depois do núcleo. Na língua Yao, por exemplo, o numeral pode assumir função de nome, como se vê em (14), ou adjetivo, exemplificado em (15). Segundo Ngunga (2004), na língua Yao, quando o numeral possui função de nome, ele não concorda com o núcleo do SN (14), e quando o numeral possui a função de adjetivo (15), ele concorda com o núcleo.

- (14) cila wuuti n’saantu  
       cila  wuuti  n’saantu  
       MOD.  NOME  NUM. FUNÇÃO NOME  
       ‘cada cinco espingardas’
- (15) cila vaandu vatatu  
       cila  **va-**andu      **va-**tatu  
       MOD.  **PREF.-**NOME  **PREF.-**NUM. FUNÇÃO ADJ.  
       ‘cada três pessoas’

(Adaptado de NGUNGA, 2004, p. 208)

Vale salientar que o modificador *cila* possui o mesmo comportamento semântico-sintático que o pronome indefinido português *cada* (NGUNGA, 2004, p. 208), sua tradução mais acessível.

Mondlane (2012), seguindo as perspectivas de Seiler (1986), Ngunga (2002), e Ngunga e Simbine (2012), diz que, na língua Changana, há apenas nove palavras diferentes para exprimir os números, elencadas em (16), sendo que essas palavras podem corresponder a quatro bases: 5, 10, 100, 1000 e 1000000.

- (16) a. n'we: 'um'  
 b. mbirhi: 'dois'  
 c. nharhu ou rharu: 'três'  
 d. mune: 'quatro'  
 e. ntlhanu: 'cinco'  
 f. khume: 'dez'  
 g. dzana: 'cem'  
 h. khulu: 'mil'  
 i. gidi: 'milhão'

(Adaptado de MONDLANE, 2012, p.21)

Segundo Mondlane (2012, p. 21), os três primeiros números (16 a, b, c) são adjetivos, exibem prefixos de concordância com a classe nominal e são postos após o nome núcleo. Os exemplos elencados em (17) ilustram isso.

- (17) a. munhu mun'we  
       mu-  nhu          mu-  n'we  
       **PREF. CL. 1 -NOME**  **PREF. CL.1-NUM. FUNÇÃO ADJ.**  
       'uma pessoa'
- b. vanhu vambirhi  
       va-  nhu          va-  mbirhi  
       **PREF. CL. 2 -NOME**  **PREF. CL. 2-NUM. FUNÇÃO ADJ.**  
       'duas pessoas'
- c. tinyoka tinharhu  
       ti-  nyoka          ti-  nharhu  
       **PREF. CL. 10 -NOME**  **PREF. CL. 10-NUM. FUNÇÃO ADJ.**  
       'três cobras'

(Adaptado de MONDLANE, 2012, p.21)

Há ainda outros numerais cardinais que pertencem à classe dos nomes, não concordam com o núcleo, e podem ser posicionados antes ou depois do nome núcleo, ou seja, há dois nomes ligados por um extra-prefixo, um morfema que indica

concordância com o núcleo, através do mecanismo de genitivização<sup>39</sup> (NGUNGA, 2004; MONDLANE, 2012). Seguem os exemplos em (18) na língua Changana.

(18) a. tsevu wa tihomu = tihomu ta tsevu

tsevu	<b>wa</b>	ti - homu
NUM. FUNÇÃO NOME	<b>EXTRA-PREF.</b>	<b>PREF. - NOME</b>
ti - homu	<b>ta</b>	tsevu
<b>PREF. - NOME</b>	<b>EXTRA-PREF.</b>	NUM. FUNÇÃO NOME
‘seis bois’		

b. kombo wa makondlo = makondlo ya kombo

kombo	<b>wa</b>	ma - kondlo
NUM. FUNÇÃO NOME	<b>EXTRA-PREF.</b>	<b>PREF. - NOME</b>
ma - kondlo	<b>ya</b>	kombo
<b>PREF. - NOME</b>	<b>EXTRA-PREF.</b>	NUM. FUNÇÃO NOME
‘sete ratos’		

c. nhungu wa sviphongo = sviphongo sva nhungu

nhungu	<b>wa</b>	svi - phongo
NUM. FUNÇÃO NOME	<b>EXTRA-PREF.</b>	<b>PREF. - NOME</b>
svi - phongo	<b>sva</b>	nhungu
<b>PREF. - NOME</b>	<b>EXTRA-PREF.</b>	NUM. FUNÇÃO NOME
‘oito bodes’		

d. kaye wa vanhu = vanhu va kaye

kaye	<b>wa</b>	va - nhu
NUM. FUNÇÃO NOME	<b>EXTRA-PREF.</b>	<b>PREF. - NOME</b>
va - nhu	<b>va</b>	kaye
<b>PREF. - NOME</b>	<b>EXTRA-PREF.</b>	NUM. FUNÇÃO NOME
‘nove pessoas’		

(Adaptado de MONDLANE, 2012, p.22)

O kimbundu, uma das línguas do grupo bantu com maior número de falantes no Brasil, sobretudo na Bahia, nos séculos da escravização africana (LUCCHESI, 2002; PESSOA DE CASTRO, 2006), possui classes nominais relacionadas em pares, cada par indicando as formas do singular e do plural. Semelhantemente a outras línguas bantu, o acréscimo de um prefixo é que indica a que classe pertence uma palavra e estabelece a relação de concordância nos nomes, nos adjetivos e nos verbos (MONDLANE, 2012; NGUNGA, 2012).

<sup>39</sup> Genitivização é o processo pelo qual dois nomes, um núcleo e um modificador, formam e qualificam um SN. Geralmente o nome e o modificador são ligados por um extra-prefixo que marca a concordância e a ordem dos constituintes, ficando facultativa ao falante a ordem, em determinadas línguas como a Changana. A genitivização é o mais produtivo mecanismo de qualificação nas línguas bantu porque pode indicar qualidade, posse, matéria ou origem (cf. NGUNGA, 2004, p. 203).

Nos exemplos (19) e (20) do kimbundo, podemos ver sentenças com SNs no singular e no plural, nas quais podemos observar que a marcação de número é estabelecida por um prefixo adjunto ao nome núcleo: o prefixo *mu-* que indica singular tanto em (19a) como em (20a); e os prefixos *ah-* e *mi-*, indicando plural e expressando a relação de concordância nos SNs de (19b) e (20b), nesta ordem.

(19) a. muhatu walamba makunde  
 mu- hatu walamba ma-kunde  
**SING. CL 1-** NOME VERBO **PL. CL. 6-** NOME  
 'A mulher cozinha feijões'

b. ahatu alamba makunde  
 a- hatu alamba ma-kunde  
**PL. CL. 2-** NOME VERBO **PL. CL. 6-** NOME  
 'As mulheres cozinham feijões'

(20) a. muzangala walamba makunde  
 mu- zangala alamba ma-kunde  
**SING. CL. 1-** NOME VERBO **PL. CL. 6-** NOME  
 'O rapaz cozinha feijões'

b. mizangala yalamba makunde  
 mi- zangala yalamba ma-kunde  
**PL. CL1-** NOME VERBO **PL. CL 6-** NOME  
 'Os rapazes cozinham feijões'

Assim como em outras línguas bantu, no kimbundo, a CN pode ocorrer pela inserção de um morfema prefixo no nome núcleo, como nos exemplos em (19) e em (20), e também pelo mecanismo de genitivização – inserção de um extra-prefixo ligando o núcleo e o modificador, como podemos ver em (21). É interessante observar que o genitivo no kimbundo estabelece uma relação de concordância dentro do SN, diferentemente do português, conforme a tradução literal mostra em (21). Enquanto no kimbundo a concordância ocorre inclusive no genitivo, no português a preposição é um elemento de entrave porque se trata de uma categoria que não se flexiona. De alguma forma, esse aspecto pode enfraquecer a marcação em todos os constituintes no PB.

(21) a. abika a kuku  
 a- bika a kuku  
**PL. CL 1-** NOME **EXTRA-PREF. CL 1** **PL.POSSE.** NOME  
 escravos avô  
 'os escravos do avô'

b. Inama ia muhatu



i- nama                      ia                      muhatu  
**PL. CL. 2- NOME EXTRA-PREF. CL 2 PL.POSSE NOME**  
 pernas    mulher  
 'as pernas da mulher'

(Adaptado de CHATELAIN, 1888-89, p.10)

De modo semelhante a outras línguas bantu, o kimbundu pode ter no SN a presença de numerais. Chatelain (1888-89) descreve o comportamento dos numerais no SN afirmando que, a depender do numeral, a concordância pode ocorrer com a inserção de um prefixo de classe para concordar com o nome núcleo ou via processo de genitivização. O Quadro 15, adaptado de Chatelain (1888-89, p.19) exhibe os numerais cardinais de 1 a 10 e exemplos de suas ocorrências em SNs.

Quadro 15. Numerais cardinais de 1 a 10, no kimbundo, e SNs correspondentes.

<b>Numeral 'tradução'</b>	<b>Exemplos de SNs 'tradução'</b>
moxi 'um'	mutu umoxi 'uma pessoa'
iari 'dois'	atu kiiari 'duas pessoas'
tatu 'três'	atu kitatu 'três pessoas'
uana 'quatro'	atu kiuana 'quatro pessoas'
tanu 'cinco'	atu kitanu 'cinco pessoas'
samanu 'seis'	atu kisamanu 'seis pessoas'
(ri) sambuari 'sete'	sambuari ria atu 'sete pessoas'
(ri) nake 'oito'	nake ria atu 'oito pessoas'
(r) ivua 'nove'	'vua ria atu 'nove pessoas'
(ri) kuinii 'dez'	kuinii ria atu 'dez pessoas'

Fonte: adaptado de CHATELAIN (1888-89, p. 19).

Os exemplos dos SNs no Quadro 15 são formados por um numeral e por um nome da classe 1. Segundo Chatelain (1888-89), o número um é regido pelo nome núcleo e a concordância é estabelecida com a inserção do prefixo **u-**. Aos números dois, três, quatro e cinco são acrescentados um prefixo da classe 3, sendo caracterizados como nomes numerais e podem ser posicionados antes ou depois do nome núcleo. Neste caso, o nome numeral e o nome núcleo são ligados por um genitivo, como podemos ver em (22).

- (22) atu kitatu        ou        kitatu kia atu  
           atu                kitatu  
 NOME NÚCLEO NOME NUM.  
 kitatu        **kia**        atu  
 NOME NUM. **GENITIVO** NOME NÚCLEO  
 'três pessoas'<sup>40</sup>

(Adaptado de CHATELAIN, 1888-89, p.19)

O número seis, *samanu*, pode ser usado de três formas: sem prefixo, para todas as classes (23); como nome numeral antes ou depois do nome núcleo (24); ou concordando com nome núcleo recebendo um prefixo de concordância de classe e número (25).

- (23) atu        samanu  
 NOME CL. 1 NUM.  
 'seis pessoas'

- (24) **Ki-**        samanu  
**PREF. DE CL. – NUM.** NOME  
 'seis'

- (25) atu kisamanu    ou        kisamanu kia atu  
           atu                kisamanu  
 NOME NÚCLEO NOME NUM.  
 kisamanu **kia**        atu  
 NOME NUM. **GENITIVO** NOME NÚCLEO  
 'seis pessoas'

(Adaptado de CHATELAIN, 1888-89, p.19)

Ainda de acordo com Chatelain, os números sete, oito, nove e dez são, na verdade, nomes da classe IV e por isso podem perder seu prefixo *ri-*, quando este aparece no genitivo. O autor destaca que oito, nove e dez sempre precedem o nome núcleo, de qualquer classe, e são ligados a ele por um genitivo, como podemos ver em (26).

- (26) nake        **ria**        mindele  
 NOME NUM. **GENITIVO** NOME NÚCLEO  
 'oito homens brancos'

<sup>40</sup> Também pode ser interpretado como 'uma trindade de pessoas' (cf. CHATELAIN, 1888-89, p.19)

Com essa breve descrição linguística da expressão de plural e de CN nos SNs das línguas bantu, podemos perceber que são tipologicamente distintas do português. Somados a isso, estão os processos sócio-históricos de formação da sociedade brasileira e do PB em contexto multiétnico, o que nos leva a considerar o contato linguístico como um elemento fundamental para a análise sociolinguística da CN no português popular da Bahia, temática a que se dispõe esta pesquisa.

## 6.2 INFERÊNCIAS SOBRE O PAPEL DAS LÍNGUAS BANTU NA CONCORDÂNCIA NOMINAL DO PORTUGUÊS RURAL DA BAHIA

Conforme descrito na seção 6.1 acerca das principais características gramaticais e linguísticas das línguas bantu, sobretudo do kimbundo, o plural dos nomes é designado através do processo de prefixação. Quanto à estrutura do SN nas línguas bantu, os constituintes modificadores são mais posicionados à direita do nome núcleo e a marca de plural é centralizada no nome e no modificador com os prefixos correspondentes às classes. Estruturas de SN com presença de numerais, que são elementos semanticamente plurais, podem ocorrer com esse constituinte à esquerda ou à direita do nome núcleo a depender da sua função no sintagma, nome ou adjetivo.

No Quadro 16, estão elencadas, de um lado, as características das línguas bantu relevantes para a proposta de análise deste trabalho, e do outro, as características do português que serviu de modelo para aquisição do português como L2 pelos africanos trazidos para o Brasil.

Quadro 16. Características das línguas bantu e do português.

<b>Características das línguas bantu</b>	<b>Características do português</b>
Aglutinantes	Flexional
Marcação de plural sincrética à de classe prefixada no nome	Marcação de plural sufixada ao nome
Concordância variável, a depender da classe do elemento não nuclear <sup>41</sup>	CN entre os elementos do SN, o que acarreta redundância

<sup>41</sup> A redundância de marcas de plural no SN das línguas bantu é variável, pois há momentos em que flexiona momentos em que não CN. Nas línguas bantu, as marcas de plural se espraiam para o SV (cf. NUNGA, 2004).

Não possuem artigo	Artigo definido átono
O nome em primeira posição é necessariamente flexionado	Quaisquer elementos flexionáveis do SN recebem a marca de plural.

Fonte: elaborado pela autora.

Podemos ver que se trata de um quadro de línguas tipologicamente diferentes – as línguas bantu e o português – que estiveram em contato intenso por um período de tempo considerável nos territórios africanos e brasileiro em condições de domínio, escravização e imposição de elementos socioculturais por parte dos portugueses aos demais povos.

Tendo em vista o conjunto de características do Quadro 16 e considerando a amplitude de respostas que uma situação de contato linguístico pode gerar, evidencio e reforço a hipótese das línguas bantu, sobretudo o kimbundu, terem fixado traços aglutinantes no português. Ressalto o fato de, na sócio-história e na formação da população do interior do Estado da Bahia, na região onde situam-se os municípios de Poções e de Santo Antônio de Jesus, ter uma significativa presença de angolanos e moçambicanos escravizados (SILVA, 2005), em sua maioria falantes de kimbundu, quicongo e umbundu e de algumas línguas do grupo kuá que, segundo Guy (2005), também marcam o plural na primeira posição do SN, além de não se observar CN.

Nas línguas faladas pela maioria dos escravos que vieram para o Brasil, encontramos dois grupos: um proveniente da área da atual Nigéria, compreendendo as línguas kuá (por exemplo iorubá e ibo), e outro consistindo nas línguas bantu da África meridional (por exemplo, kimbundu e kikongo). Os dois grupos são diferenciados por várias características lingüísticas, mas têm uma característica que é mais ou menos geral em todas essas línguas: o marcador de plural normalmente ocorre na primeira posição do sintagma nominal. No grupo kuá, por exemplo, não há concordância, e na maioria das línguas a indicação explícita do número no sintagma nominal é opcional: qualquer sintagma sem marcador pode ser interpretado como singular ou plural, dependendo do contexto. Mas quando é preciso explicitar o fato de um sintagma ser plural, há uma palavra que é colocada no início do sintagma para indicar isso (GUY, 2005, p. 28).

Proponho que, respondendo a segunda questão proposta no início deste Capítulo, nos primeiros momentos do contato, tenha ocorrido a aquisição do artigo, uma categoria funcional e átona, em adjunção à margem esquerda da raiz nominal, retomando o padrão das línguas bantu, que são aglutinantes, elementos funcionais são prefixados.

Sendo o português uma língua que estabelece a concordância entre todos os seus elementos, não seria diferente com o artigo. Dessa forma, o artigo, inexistente nas línguas bantu, foi adjungido à raiz marcando singular ou plural ao nome, mantendo o padrão dessas línguas. Isso é possível, pois a maior parte dos SNs nas línguas são constituídos de apenas dois elementos Det + Nome, acarretando a robustez das evidências.

Posteriormente, os aprendentes do português, a partir de novos *inputs*, reinterpretam o Det como uma categoria independente da raiz, mas que tem escopo sobre ela no que diz respeito sobre a marcação de plural, estabelecendo o padrão mais comum do PB, que é a preferência pela marcação de plural apenas no determinante ou na primeira palavra do SN.

Quanto à aquisição da regra de concordância, como estudos sobre o tema tem demonstrado ser um comportamento entre os mais jovens, proponho que a escolarização e outros fatores sociais que levaram ao contato com normas de prestígio, a aquisição se difunde a partir de elementos que têm valor semântico de número, como os numerais e os quantificadores indefinidos, os quais reforçam a ideia de plural do SN.

Para substanciar a hipótese proposta, retomo a hipótese de Aboh (2015). Suas pesquisas tiveram como base os estudos de Mufwene (2001, 2008) sobre competição e seleção de traços num sistema híbrido em que características sintáticas e semânticas de diferentes línguas em situação de contato podem se combinar para formar uma nova língua. Nesse sentido, os aprendentes teriam a sua disposição diversas possibilidades de pronúncias, alternativas lexicais, estruturas gramaticais das línguas em contato, que constituem a formação de um banco de traços ou piscina de traços (*feature pool*).

Parece que, na aquisição do português, o traço de marcação de plural é mais evidente que o traço de classe, visto que no português os nomes não são categorizados em classe e no SN a marca de plural ocorre em todos elementos de forma redundante. Sendo as línguas bantu uma língua aglutinante, à raiz nominal são adjungidas todas as informações gramaticais necessárias.

A posição mais à direita da marca de plural pode ser interpretada como sendo uma forte concorrente na aquisição da L2 desses indivíduos. Outro traço que parece ser relevante é o fato de que elementos funcionais são aglutinados à margem esquerda, o que teria levado a aglutinação do artigo, que expressa a referencialidade

do nome, às raízes nominais. Dessa forma obtém-se a mesma estrutura do kimbundu, nomes carregam marca de plural na margem esquerda: *asmenina*. É possível notar que o caráter átono do artigo definido contribui para essa interpretação.

Esses SNs deveriam ser robustos no momento de aquisição. Neste trabalho, os SNs com essa estrutura são os mais frequentes com 2.271 ocorrências, correspondendo a cerca de 47% do total dos dados. Estabelecido o local de marcação do plural, a concordância é pouco significativa, considerando a redundância. Segundo Aboh (2015), os falantes tendem a interpretar a estrutura da L2 conforme a estrutura da L1 e, nesse processo, descartam traços com informações percebidas como redundantes. Posteriormente, evidências da possibilidade de introdução de itens lexicais entre artigo e nome ,- consolidam o artigo como local da marcação de plural nesses dialetos.

Outro aspecto que explicaria a não marcação do plural em todos os elementos é o fato de que a concordância entre os elementos do SN nas línguas bantu depende da categoria do constituinte, principalmente dos numerais. Dessa forma, a marcação em todos os elementos pode não ter sido traço forte para a aquisição da CN entre todos os elementos.

Por fim, considerando, conforme Capítulo 5, a aplicação da regra de concordância nessas comunidades está em processo aquisicional e se difunde a partir dos sintagmas cujo núcleo é um numeral ou um quantificador. Essa proposta também justifica o motivo da variável saliência ter um comportamento diferente da escala proposta por Scherre (1988) neste e nos trabalhos sobre o português rural afro-brasileiro (ANDRADE, 2006), além de justificar porque o princípio da simplicidade não se aplica, afinal teríamos apenas um item lexical que carrega marca de plural, não sendo pertinente falar em concordância.

Essas hipóteses explicativas responderiam, portanto, as questões levantadas no início dessa seção, dedicada a levantar algumas proposições sociolinguísticas a respeito do comportamento da CN no português popular da Bahia.

Como vimos na seção 6.1, no kimbundu, os substantivos agrupam-se num certo número de morfemas que se colocam no seu início para indicar se se trata de singular ou plural: são os prefixos nominais ou classificadores. Por esta razão, os falantes de quimbundo que têm o português como L2 tendem a aproximar as duas línguas, pela introdução das regras do kimbundo. Vejamos os dados em (27) e (28) do trabalho de António e Osório (2018), no qual os autores estudaram o português

falado na região do Kwanza Norte, Angola, em que os falantes são bilíngues em kimbundu e português.

- (27) a. Ele tem **muitos carros** (português padrão)  
 b. Ele tem **muito scarro** (português do Kwanza Norte)
- (28) a. Temos **muitas casas** (português padrão)  
 b. Temos **muita scasa** (português do Kwanza Norte)

(ANTÓNIO e OSÓRIO, 2018, p. 3294)

Nesses exemplos, podemos observar que os falantes de kimbundu e de português como L2 fazem construções sintagmáticas na tentativa de aproximar as duas línguas, de maneira que a variedade reestruturada do português apresenta o apagamento da marca de plural no primeiro elemento do sintagma e a inserção da marca de plural como um elemento prefixal no nome núcleo. Essas construções evidenciam que as regras do kimbundu são transpostas para o português.

Corroboro, então, com o que propõem Mufwene (2001, 2008), Aboh (2015) e Aboh e Ansaldo (2017) no que tange às características das estruturas que compõem a variedade de língua originada de situações de intenso contato de línguas de tipologias diferentes: não se trata de uma mera transferência gramatical de alguma estrutura de uma das línguas em contato para a língua emergente ou de uma simplificação morfológica, tendo em vista que os dados linguísticos e a sócio-história apontam para uma adequação da estrutura ao contexto da língua emergente de acordo com a atribuição de pesos dos traços integrantes do banco, independentemente o peso que possuíam em suas línguas de origem.

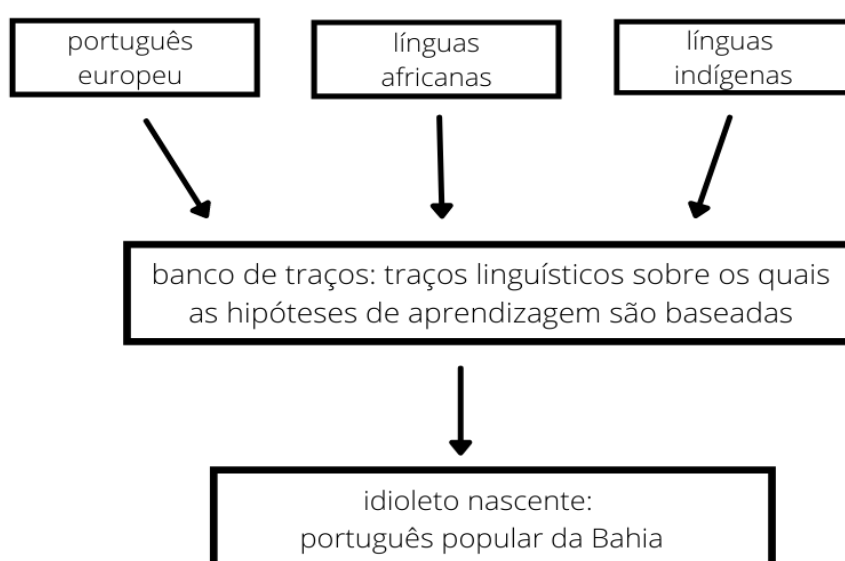
A CN pode ser um exemplo disso, já que o peso atribuído à marcação de plural no SN pelos aprendentes de português L2, apenas no primeiro elemento do sintagma, apresenta um peso diferente no PE e na norma padrão do PB, em que se preza a marcação em todos os constituintes do SN. E fora justamente a variedade linguística do português adquirido pelos aprendentes adultos de L2 que se tornou a língua materna das gerações seguintes no interior da Bahia, o português popular, que carrega marcas socialmente estigmatizadas, como a ausência de marcação de plural em todo o SN.

Levando em conta o que Aboh (2015) propôs sobre o processo de aquisição da língua alvo em contextos de contato linguístico intenso não ser resultado de uma replicação imperfeita do input, aponto algumas proposições com base nos resultados quantitativos, qualitativos e probabilísticos obtidos neste trabalho.

A partir da competição e seleção de traços, surgiriam idioletos, nos quais alguns traços seriam variantes linguísticas dominantes a depender do contexto social do processo de aquisição, sendo que posteriormente, as variantes mais predominantes nos idioletos fariam parte da comunidade de fala. Assim, as variantes possuem pesos diferentes, tendo preferência, pelos falantes, as variantes menos marcadas (MUFWENE, 2008, p. 115-117).

Mufwene (2008) ainda ressalta que, quando o banco de traços linguísticos é composto por traços de línguas muito distintas, a exemplo do PE e das línguas bantu, ocorre uma alteração no equilíbrio de poder (*balance of power*), isto é, uma mudança em relação ao peso atribuído às variantes, de forma que a seleção de traços contribui para a formação de uma nova língua (cf. MUFWENE, 2008, p. 32; 118; 123). Adaptando a figura de Aboh (2015, p. 115) para o contexto do contato linguístico ocorrido na formação do PB, ilustro através da Figura 3 como teria se formado o banco de traços que contribuiriam para a formação do português popular da Bahia.

Figura 3. Representação de banco de traços linguísticos.



Fonte: elaborada pela autora com base em Aboh (2015, p. 115).



Ampliando a hipótese de Aboh (2015), Aboh e Ansaldo (2017) indicam quais traços linguísticos são mais evidentes e mais competitivos em relação a outros traços. Segundo os autores, os traços com características semânticas como tópico, foco, especificidade etc. são os que mais entram em competição, por outro lado, os traços mais funcionais, como pessoa e número, são menos competitivos, principalmente se se apresentarem redundantes. Essa proposição pode ser observada através dos resultados do processamento dos dados da presente pesquisa, pois constatamos que a presença dos numerais e de quantificadores indefinidos, que são elementos com transparência semântica de plural, e os dados apontam para o fato de esses serem os contextos de aquisição da regra, com chances maiores em 2.80 e 1.91 *odds*, respectivamente.

Assim, assumo que a CN no português popular do interior da Bahia pode exibir o hibridismo resultante do processo de seleção e competição de traços que ocorre a partir da formação de um banco de dados – *feature pool* (MUFWENE, 2001, 2008) – impulsionado pelo contato linguístico intenso do português com as línguas bantu, principalmente. Destaco ainda que as proposições expostas nesta seção possuem caráter hipotético, visto que, a constituição da sociolinguística do PB é complexa e não se pode categorizar com rjeza todas as propriedades dos idioletos de europeus, africanos, indígenas, dentre outros, envolvidos na em sua formação.

### 6.3 SÍNTESE

Neste Capítulo, fiz uma breve descrição da marcação de plural das línguas bantu, as quais foram bastante relevantes para a formação do PB, sobretudo do português da Bahia. Associados os aspectos sócio-históricos e sociolinguísticos ao comportamento morfossintático e semântico da CN nas línguas bantu e no português, elenquei algumas diferenças tipológicas dessas línguas para tecer considerações sobre os resultados desta pesquisa.

Considerando o contato intenso entre línguas de tipologias distintas, testei a hipótese levantada por Guy (2005) sobre a tendência de o PB popular marcar o primeiro elemento do SN ser um efeito da interferência das línguas bantu e, com base, nos pressupostos da competição e seleção de traços (Mufwene, 2001, 2008) e a teoria nula (Aboh, 2015), propus que teria ocorrido um processo de reanálise

gramatical em que o morfema indicador de plural nos nomes núcleos das línguas bantu, os prefixos, foram reinterpretados de maneira que elementos posicionados à esquerda do núcleo, preferencialmente, receberiam a marca morfológica de plural -s.

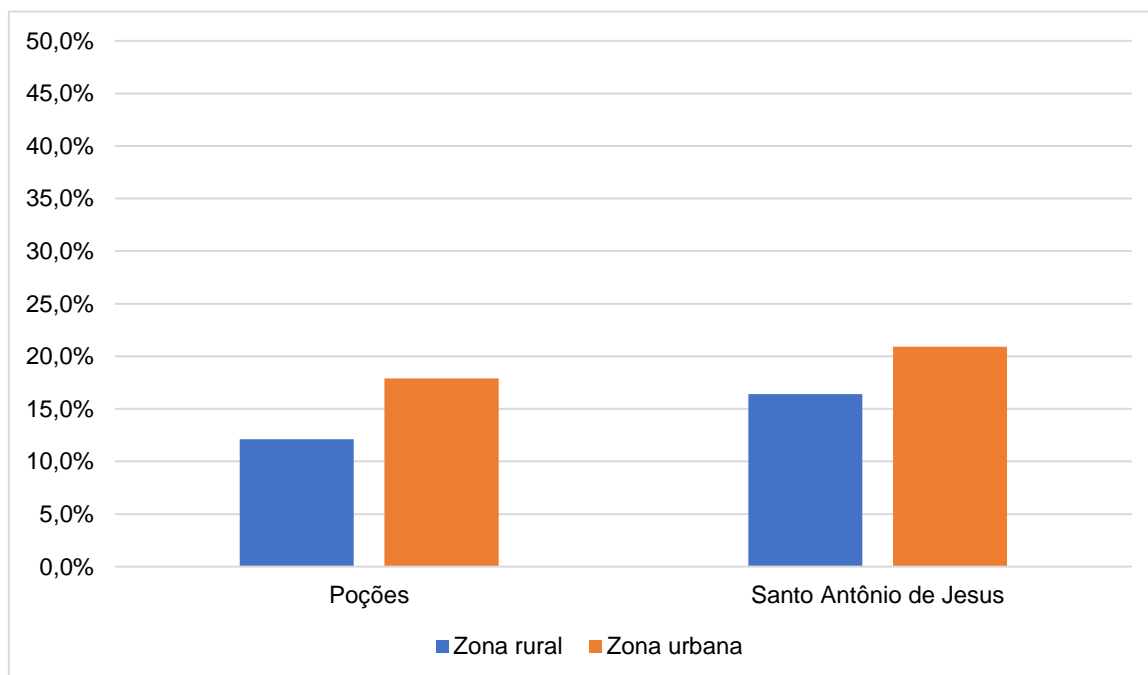
Por fim, teço a inferência de que a CN no português popular do interior da Bahia parece apresentar o hibridismo resultante do processo de seleção e competição de traços (MUFWENE, 2001, 2008; ABOH, 2015), tendo como força estimuladora o contato linguístico intenso do português com as línguas bantu.

## 7. CONSIDERAÇÕES FINAIS

Neste trabalho, analisei o fenômeno variável da CN no âmbito do SN no português popular do interior da Bahia com o intuito principal de contribuir para a discussão sobre essa variedade do português, considerando o contato linguístico intenso como aspecto formador fundamental de suas particularidades. Os demais objetivos desta pesquisa giraram em torno de descrever a regra variável da CN, tendo em vista verificar o *status* desse fenômeno num *continuum* linguístico que tem, num extremo, as comunidades rurais afro-brasileiras e no outro, a capital do estado, Salvador.

Para pormenorizar a verificação do *continuum*, destaco, no Gráfico 6, novamente os resultados da distribuição da aplicação da regra de CN no eixo rural-urbano nos municípios de Poções e de Santo Antônio de Jesus, conforme os resultados descritos no Capítulo 5: zona rural de Poções: 12,1%; zona urbana de Poções: 17,9%; zona rural de Santo Antônio de Jesus: 16,4%; zona urbana de Santo Antônio de Jesus: 20,9%.

Gráfico 6. Aplicação da regra de CN nas zonas rurais e urbanas de Poções e de Santo Antônio de Jesus.

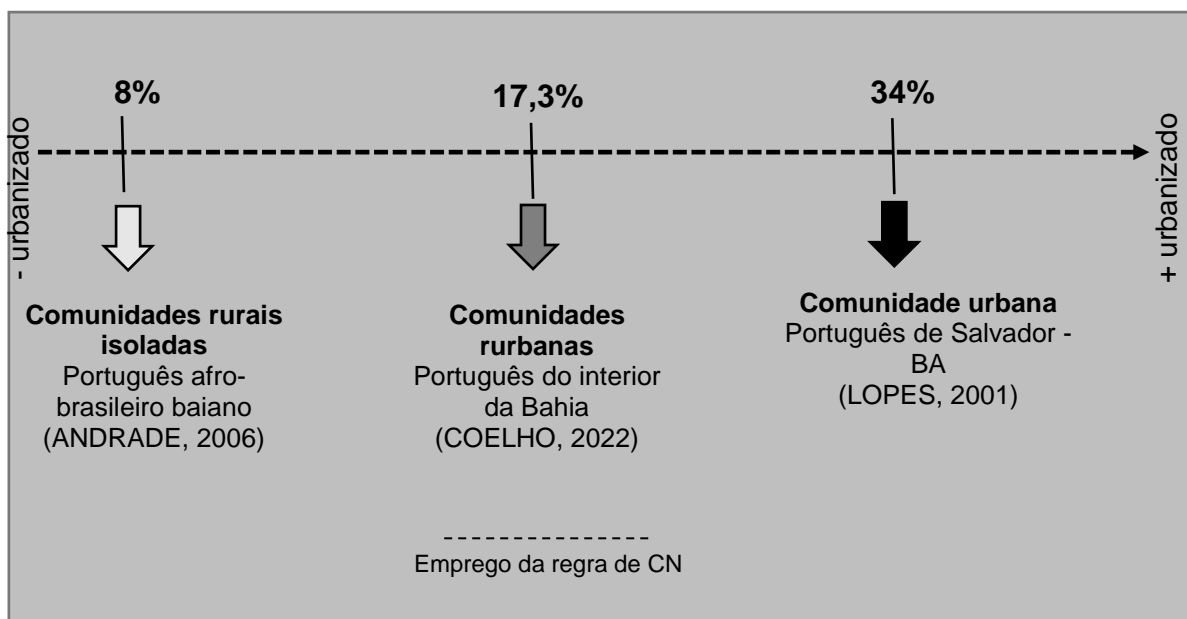


Fonte: elaborado pela autora.

Os resultados ilustrados no Gráfico 6 evidenciam os efeitos da urbanização na variedade linguística em estudo mediante seu processo de constituição histórica, conforme características elencadas por Bortoni-Ricardo (1998; 2005) e Lucchesi (2015). Reitero o que fora descrito no Capítulo 4 sobre a sócio-história de Poções e de Santo Antônio de Jesus, na qual o cenário que movimenta a economia dessas comunidades pode ser um aspecto importante para a compreensão da dialeção da CN na região. Poções é um município predominantemente rural, cujas atividades socioeconômicas giram em torno da agricultura. Santo Antônio de Jesus, por sua vez, se destaca pelo comércio, pelo setor de serviços, pela indústria de móveis bastante conhecida desde as regiões vizinhas até a capital do estado. Nesse panorama, os falantes de Poções estabelecem relações sociais de alta densidade e multiplex mais distantes do efeito das normas urbanas cultas, empregando menos a regra de CN. Já os falantes de Santo Antônio de Jesus empregam mais a CN por estabelecerem relações de baixa densidade e uniplex devido ao maior índice de urbanização e proximidade da capital.

A respeito da verificação do *status* da CN no *continuum* linguístico no Estado da Bahia, parece que esse fenômeno, apesar de variar em todos os dialetos do PB, tem uma tendência à aplicação da regra na proporção de aproximação dos grandes centros urbanos. A Figura 4 ilustra o comportamento proporcional da CN no *continuum* sociolinguístico nas comunidades rurais afro-brasileiras baianas (ANDRADE, 2006), no português do interior, segundo dados obtidos neste trabalho, e na cidade de Salvador (LOPES, 2001). A verificação da CN nesse *continuum* indica que o português popular do interior da Bahia possui as características de variedade urbana (BORTONI-RICARDO, 1998, 2005; LUCCHESI, 2015) conforme explanação no Capítulo 4, e corrobora a perspectiva de Lucchesi (2015) sobre a existência de uma polarização linguística e social do PB.

Figura 4. A CN no *continuum* sociolinguístico no português da Bahia.



Fonte: elaborada pela autora.

Após a apresentação do fenômeno da CN, no Capítulo 2, fiz uma sistematização dos estudos que forneceram matrizes teórico-metodológicas de relevância para este trabalho (SCHERRE, 1988; LOPES, 2001; ANDRADE, 2003, 2006; BAXTER 2009). Observei que o comportamento sociolinguístico da CN pode variar, sobretudo, de acordo com as normas linguísticas cultas e populares do PB. Essas normas se diferenciam diatopicamente porque foram formadas a partir de situações sócio-históricas distintas. Nos centros urbanos, predominou a variedade do português mais próxima ao PE supervalorizada e reproduzida por uma pequena elite com acesso aos níveis mais altos de escolaridade. Nas regiões mais interioranas, onde índios, africanos e seus descendentes se fixaram em maior quantidade, o português predominante foi aquele mais distante do modelo europeu.

Nesse cenário, a presença de CN tende a ter mais frequência na fala de residentes das zonas urbanas (SCHERRE, 1988; LOPES, 2001, BRANDÃO e VIEIRA, 2012; LUCCHESI e DÁLIA, 2020) do que entre os falantes das zonas rurais (ANDRADE, 2003, 2006; BAXTER, 2009, LUCCHESI e DÁLIA, 2020). Esse panorama reforça a proposição de que o português do interior da Bahia apresenta evidências de uma variedade rurbarana.

Compreendendo que a realidade sociolinguística do Brasil é complexa, conforme apresentei brevemente nos Capítulos 3 e 4, analisei dados da fala de 48 informantes (24 do município de Poções e 24 de Santo Antônio de Jesus) dos sexos

masculino e feminino, divididos em três faixas etárias: jovem – 25 a 35 anos; intermediária – 45 a 55 anos; idosa – acima dos 65 anos. Utilizei o modelo teórico-metodológico da Sociolinguística Variacionista (Labov,1972), exibido no Capítulo 4, para a análise descritiva da CN no português popular da Bahia, que constituiu parte do Capítulo 5. Neste também empreendi a análise quantitativa dos dados, para a qual recorri ao R com a interface RStudio, programa computacional utilizado nas mais diversas áreas da ciência para análises estatísticas.

A análise dos dados foi realizada a partir da perspectiva sintagmática, na qual todo o SN é uma unidade de análise. A hipótese geral da pesquisa partiu do pressuposto de que a variação no uso da regra de CN, no português popular do interior da Bahia, seria motivada por fatores linguísticos e extralinguísticos que refletem características do resultado de contato linguístico do português com as línguas africanas. A quantificação geral dos dados apontou uma proporção de 17,3% de aplicação da regra de concordância nas 4.867 ocorrências do fenômeno no *corpus*, confirmando a hipótese geral de que haveria menos dados com concordância do que com concordância.

Após a verificação da distribuição dos dados, feitas as devidas modificações, apliquei testes estatísticos para a análise univariada (qui-quadrado) e para a análise multivariada (regressão logística com efeitos mistos) para verificar a relevância significativa de cada variável independente em relação à variável dependente – presença *versus* ausência de CN –, bem como a relevância das variáveis independentes quando combinadas entre si.

Os resultados indicaram que as variáveis estaticamente mais relevantes para a variação da CN no português popular do interior da Bahia foram: configuração sintagmática do SN, número de constituintes flexionáveis do SN, saliência fônica, sexo, faixa etária e escolaridade. De maneira geral, há semelhanças entre os resultados desta pesquisa com os trabalhos de Scherre (1988) para o Rio de Janeiro, Lopes (2001) para Salvador, Andrade (2003, 2006) e Baxter (2009) para as comunidades afrodescendentes do interior da Bahia, Helvécia, Cinzento, Rio de Contas e Sapé. No entanto, a análise estatística de algumas variáveis clássicas dos estudos de concordância, a configuração sintagmática do SN e a saliência fônica, apresentaram resultados que podem indicar peculiaridades do português falado em Poções e em Santo Antônio de Jesus.

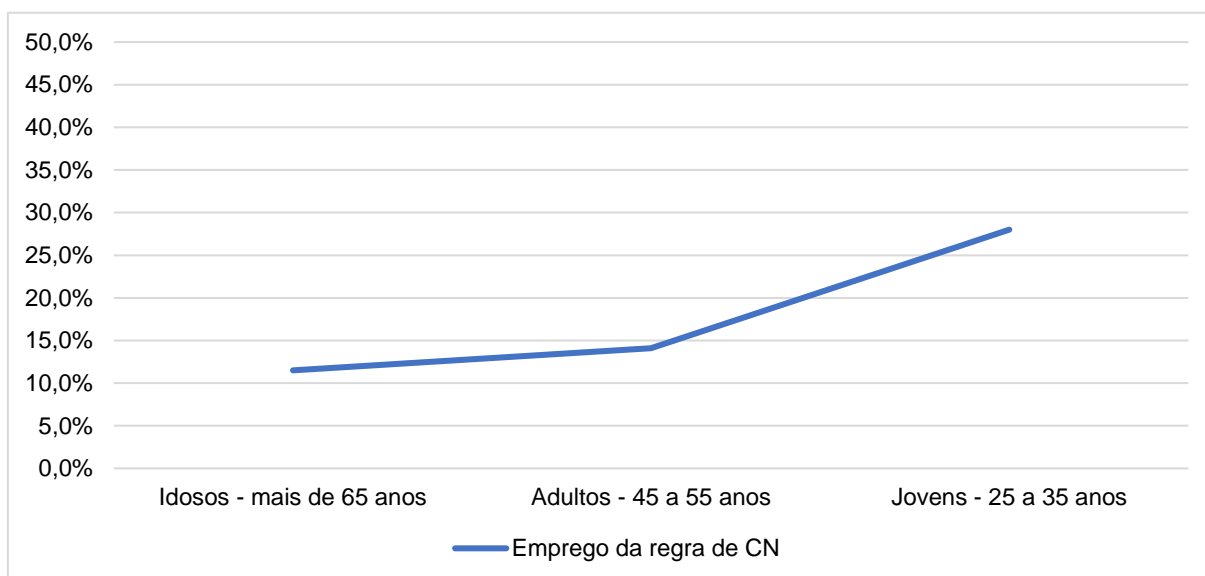
Conforme a análise qualitativa e quantitativa exibida no Capítulo 5, os resultados da configuração sintagmática do SN, também corroborados com os resultados da variável número de constituintes flexionáveis do SN, o princípio de simplicidade (LUCCHESI, 2000) parece não ser um aspecto favorecedor da aplicação da regra de CN. Essa inferência é explicada pelo fato de que é a presença de um item com valor semântico de plural, como numerais e quantificadores indefinidos, que favorece a concordância e não o fato de o SN ter uma estrutura simples, nos termos usados por Lucchesi (2000).

Quanto à saliência fônica, nos estudos sobre CN, os itens mais salientes possuem uma tendência maior à aplicação da regra de concordância. No entanto, apesar de ser estaticamente relevante, neste trabalho, essa variável, quando combinada com as demais variáveis em análise, não favoreceu a aplicação da CN. A análise qualitativa permitiu identificar possíveis expressões cristalizadas, as quais podem ter enviesado os dados.

O resultado da variável sexo indicou que, embora os falantes de ambos os sexos, feminino e masculino, apresentem pouco emprego da regra de CN, o sexo masculino desfavorece menos a CN do que o sexo feminino. Quanto à escolaridade, a análise estatística apontou uma correlação fraca com o fenômeno da CN, porém com um resultado contrário à hipótese alternativa, em que os falantes com menos escolaridade realizaram mais a CN. Esse resultado foge ao padrão é semelhante aos resultados de Andrade (2006) sobre as comunidades afro-brasileiras, nas quais a escola não possui um papel tão forte como agente de aquisição da norma culta.

No tocante ao quadro de variação da comunidade, a análise dos fatores linguísticos indicou um comportamento aquisicional da regra de CN que parece se difundir no SN em que o nome é precedido de um numeral ou quantificador. Junto a esse indício, o resultado da variável social faixa etária parece corroborar a tendência de aquisição da norma de prestígio: os falantes mais jovens aplicam mais a regra de CN do que os falantes mais velhos, conforme apresenta o Gráfico 7.

Gráfico 7. Distribuição da proporção de aplicação da regra de CN no português popular da Bahia quanto à faixa etária.



Fonte: elaborado pela autora.

Com a realização deste trabalho, propus uma observação da CN a partir da perspectiva da variação e mudança das línguas impulsionadas pelos efeitos do contato linguístico considerando o processo de competição e seleção de traços (MUFWENE, 2001, 2008) e à teoria nula (ABOH, 2015; ABOH e DEGRAFF, 2017). No Capítulo 6, defendi a tese de que, diante do contexto multilíngue de contato intenso na formação do português baiano, teria ocorrido um processo de reanálise gramatical na aquisição do português como L2.

Nesse processo, o morfema indicador de classe – e plural - nos nomes núcleos das línguas bantu, os prefixos, foram reinterpretados de maneira que elementos posicionados à esquerda do núcleo, preferencialmente, receberiam a marca morfológica de plural -s, não apenas se mantendo no primeiro elemento do sintagma (GUY, 2005), mas também seguindo a manutenção inicial da gramática de aglutinação dos elementos funcionais. A proposta de tese que apresento encontra similaridade no que os autores António e Osório (2018) propõem para o português falado na região do Kwnaza Norte, Angola, onde os falantes de kimbundu e de português como L2 fazem construções sintagmáticas reestruturadas do português com o apagamento da marca de plural no primeiro elemento do sintagma e a inserção da marca de plural como um elemento prefixal no nome núcleo (cf. seção 6.2 do Capítulo 6 deste trabalho).



Tendo em vista que as línguas bantu são línguas aglutinantes, é provável que a concordância tenha sido um traço não tão evidente no processo de combinação e recombinação de traços, além de se apresentar como redundante (ABOH e DEGRAFF, 2017) a ponto de não ser adquirida como necessária em todos os itens constituintes do sintagma, como se apresenta no PE e na variedade culta do PB. Outra característica tipológica que explicaria a não marcação do plural em todos os elementos é o fato de a concordância entre os elementos do SN nas línguas bantu depender da categoria do constituinte, sobretudo dos numerais. Nesse sentido, a marcação em todos os elementos pode não ter sido um traço robusto para a aquisição da CN em todos os elementos do sintagma.

Diante dos resultados deste trabalho, reitero a hipótese de que o português do interior da Bahia teria se formado sob condições sócio-históricas de contato linguístico e cultural de maneira que suas peculiaridades linguísticas se configurariam não como réplica das línguas africanas ou simplificação do PE, mas sim como uma nova variedade de língua (CHAUDENSON, 1992, 2001; MUFWENE, 2001, 2002, 2008).

Por fim, ressalto que esta pesquisa abre a possibilidade de novos caminhos para os estudos sociolinguísticos da CN, bem como do português baiano, haja vista a necessidade de investigar o uso do quantificador *tudo*, conforme apresentei na subseção 5.2.1.1. Deixo em aberto as seguintes questões: o uso peculiar do quantificador *tudo* seria uma reanálise de alguma marca de plural das línguas bantu? Quais os contextos linguísticos e extralinguísticos que levam os falantes a utilizarem o *tudo* ora como advérbio de intensidade ora como quantificador?

Essas questões têm como objetivo provocar o desdobramento de trabalhos vindouros e, conseqüentemente, a ampliação do conhecimento da língua portuguesa e suas variedades, valorizando e respeitando a sócio-história, a cultura e a realidade da formação de cada uma delas.

## REFERÊNCIAS

- ABNEY, S. 1987. **The English noun phrase in its sentential aspect**. Tese de doutorado. Massachusetts Institute of Technology, Cambridge.
- ABOH, E. 2009. Competition and Selection. That's all. In: ABOH, E.; SMITH, N. (eds.). **Complex processes new languages**. Amsterdam Philadelphia: J. Benjamins.
- ABOH, E. 2015. **The emergence of hybrid grammars**: language contact and change. Cambridge: Cambridge University Press.
- ABOH, E.; ANSALDO, U. 2007. The role of typology in language creation: A descriptive take. In: ANSALDO, U.; MATTHEWS, S.; LIM, L. (Ed.) **Deconstructing Creole**. Amsterdam: John Benjamins, p. 39-56.
- ABOH, E.; MATTHEWS. 2007. Deconstructing creole: The rationale. In: ANSALDO, U.; MATTHEWS, S.; LIM, L. (Ed.) **Deconstructing Creole**. Amsterdam: John Benjamins, p. 1-14.
- ABOH, E.; DEGRAFF, M. 2017. A Null Theory of Creole Formation. In: ROBERTS, I (ed). **The Oxford Handbook of Universal Grammar**. Oxford Press, p. 1-67.
- ALENCASTRO, L. F. 2000. **O trato dos viventes** – formação do Brasil no Atlântico sul. São Paulo: Companhia das Letras.
- ALMEIDA, A. P. 2006. **A concordância verbal na comunidade de São Miguel dos Pretos, Restinga Seca, RS**. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul.
- ALMEIDA, K. R. 2013. Rede urbana, cidades e municípios pequenos no Brasil: um estudo sobre as possibilidades de planejamento cooperado em Poções, Caetanos e Bom Jesus da Serra. In: **XIV Encontro de Geógrafos da América Latina, 2013, Lima**. XIV Encontro de Geógrafos da América Latina. Disponível em: <http://observatoriogeograficoamericalatina.org.mx/egal14> Acesso em 23 dez 2021
- ANDRADE, P. R. de. 2003. **Um fragmento da constituição sócio-histórica do português do Brasil: variação na concordância nominal de número em um dialeto afro-brasileiro**. Dissertação de Mestrado. Universidade Federal da Bahia.
- ANDRADE, P. R. de. 2006. A concordância nominal de número no português afro-brasileiro. **Anais da XX Jornada – GELNE – João Pessoa-PB**. 2075-2082.
- ANTONINO, V. 2007. **A concordância nominal em predicativos do sujeito e em estruturas passivas no português popular do interior da Bahia**. Dissertação de Mestrado. Universidade Federal da Bahia.
- ANTONINO, V. 2012. **Português popular de Salvador: uma análise da concordância nominal em predicativos do sujeito e em estruturas passivas**. Tese de doutorado. Universidade Federal da Bahia.

ANTÓNIO, J. L.; OSÓRIO, P. 2018. História e variação morfossintática entre o quimbundo e o português na região de Kwanza Norte, em Angola: formação do plural, concordância de gênero, e formação do diminutivo e do aumentativo dos substantivos. **Revista Fórum Linguístico**, Florianópolis, v. 15, n. 4, p. 3289-3302.

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.

ARAUJO, S. S de F.; COUTINHO, S. 2004. O município de Santo Antônio de Jesus no Recôncavo Baiano. In: **PROJETO VERTENTES**. Disponível em: <<http://www.vertentes.ufba.br/>>. Acesso em: 11 mar. 2019

ARGOLO, W. 2013. Colonização e Língua Geral: o caso do sul da Bahia. **PAPIA**, São Paulo, 23(1), p. 75-96.

BAGNO, M. 2011. **Gramática Pedagógica do Português Brasileiro**. São Paulo: Parábola Editorial.

BAKER, P. 1994 Creativity in creole genesis. In: **Creolization and Language Change**. Dany Adone and Ingo Plag (eds.). Tübingen: Max Niemeyer. 65–84.

BAPTISTE, M. 2016. Creole formation and L2 acquisition: On re-evaluating processes and labels. **Journal of Pidgin and Creole Languages**, p. 361–389.

BARBOSA, P. A. 2013. Elementos essenciais para um entendimento dos limites e vantagens da estatística inferencial na pesquisa fonética. **ReVEL**, edição especial n. 7, p. 51-67

BAXTER, A. N. 1992. A contribuição das comunidades afro-brasileiras isoladas para o debate sobre a crioulização 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, p. 72-90.

BAXTER, A. 1996. Línguas pidgin e crioulas. In: FARIA, I. *et. al.* (Orgs). **Introdução à Linguística Geral e Portuguesa**. Lisboa: Caminho, p. 535-549.

BAXTER, A. N. 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 Pidgnização 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. 1981. **Roots of language**. Ann Arbor: Karoma Publishers, Inc.

BICKERTON, D. 1984. The language bioprogram hypothesis. **Behavioral and Brain Sciences**, 7, 173–221.

BICKERTON, D. 1988. Creole languages and the Bioprogram. In: NEWMAYER, F. (Ed.). **Linguistics: the Cambridge survey**. v.2. Cambridge: Cambridge University Press, p. 268-284.

BORGES, L. C. 2003. A língua geral: revendo margens em sua deriva. In: FREIRE, José Ribamar Bessa; ROSA, Maria Carlota (Org.). **Línguas Gerais: política lingüística e catequese na América do Sul no período colonial**. Rio de Janeiro: EdUERJ, p. 113-131.

BORTONI-RICARDO, S. M. 1998. A análise do português brasileiro em três continua: o continuum rural-urbano, o continuum de oralidade-letramento e o continuum de monitoração estilística. In: GROBE, Sybille; ZIMMERMANN, K. **“Substandard” e Mudança no Português do Brasil**. ed. Frankfurt am Main: TFM, p. 153-188.

BORTONI-RICARDO, S. M. 2005. **Nós chegemos na escola, e agora?: Sociolingüística e educação**. São Paulo: Parábola.

BRANDÃO, S. F. 2013. Patterns of plural agreement within the Noun Phrase. **Journal of Portuguese Linguistics**, 12 (2): 51-100.

BRANDÃO, S. F. 2015. Concordância nominal em três variedades do português: resultados gerais, novas indagações. **Cuadernos de la ALFAL**, n.7. 36-52.

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

CÂMARA JÚNIOR, J. M. 1981. **Dicionário de lingüística e gramática**. 10ª ed. Petrópolis: Vozes, p. 42, 227-230.

CARDIM, F. 1925. **Tratados da terra e da gente do Brasil**. Rio de Janeiro: Editores – J. Leite & Cia.

CASTILHO, A. T. de. 2010. **Nova gramática do português brasileiro**. São Paulo: Contexto.

CHAMBERS, J. M. 2020. S, R, and Data Science. **The R Journal**, v. 12, n. 1, p. 462-476. Acesso em 28 dez 2021.

CHATELAIN, H. 1888-1889. **Grammatica elementar do kimbundu ou língua de Angola**. Genebra: Typ. De Charles Schuchardt.

CHAUDENSON, R. 1992. **Des îles, des hommes, des langues: essai sur la créolisation linguistique et culturelle**. Paris, L'Harmattan Publishers.

CHAUDENSON, R. 2001. **Creolization of Language and Culture**. London: Routledge.

CHAVES, R. G. 2014. Princípio de saliência fônica: isso não soa bem. **Letrônica**, 7(2): 522-550.

CORBETT, G. 2000. **Number**. Cambridge: Cambridge University Press.

COSTA, P. R. da. 2011. **Estatística**. Santa Maria: Universidade Federal de Santa Maria. 95 p. Disponível em: <https://www.ufsm.br/app/uploads/sites> Acesso em: 13 de nov. 2021

COUTO, H. H. do. 1996. **Introdução ao estudo das línguas crioulas e pidgins**. Brasília: Editora da UNB.

D'ANGELIS, W. R. 2019. **Revitalização de línguas indígenas**: o que é? como fazemos. Editora Curt Nimuendajú: Campinas, SP.

DEGRAFF, M. 2002. Relexification: A reevaluation. **Antropological linguistics**. V. 44, n.4.

DEGRAFF, M. 2003. Against Creole exceptionalism. **Language** 79 (2), p. 391 – 410.

*ENCICLOPÉDIA DAS LÍNGUAS NO BRASIL (ELB)*. Conceitos Lingüísticos: Língua Franca. **Laboratório de Estudos Urbanos**. *Universidade Estadual de Campinas (Unicamp)* Disponível em: < <https://www.labeurb.unicamp.br/> > Acesso em 27 nov. 2019.

FARACO, C. A. 2016. **História sociopolítica da língua portuguesa**. São Paulo: Parábola Editorial.

FERNANDES, M. 1996. **Concordância nominal na região sul**. Dissertação de mestrado. Florianópolis: UFSC, 143f.

FERRO, M. 2008. **História das colonizações**: das conquistas às independências (séculos XIII a XX). São Paulo: Companhia das Letras.

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.

FIGUEIREDO, M. C. 2004. O objeto direto anafórico no dialéto rural afro-brasileiro. Dissertação de mestrado Universidade Federal da Bahia.

FRANCHETO, B. 2020. Língua(s): cosmopolíticas, micropolíticas, macropolíticas. **Campos**, vol. 21, n.1, p. 21–36. Disponível em: <https://revistas.ufpr.br/campos> Acesso em 04 de ago. 2021.

FULGÊNCIO, L. 2008. Expressões fixas e idiomatismos do português brasileiro. 2008. Tese (Doutorado em Linguística e Língua Portuguesa) – Pontifícia Universidade Católica de Minas Gerais, PUC, Belo Horizonte.

FUNAI. **Coletânea de Legislação Indigenista Brasileira**. Disponível em: [http://www.funai.gov.br/projetos/Plano\\_editorial/livro18.htm](http://www.funai.gov.br/projetos/Plano_editorial/livro18.htm) Acesso em 21 nov. 2020

GODOI, L. 2005. **A palavra *tudo* como quantificador universal puro no Português Brasileiro**. Monografia (Bacharel em Letras). Universidade Federal de Minas Gerais: Belo Horizonte.

GODOI, L.; CANÇADO, M. 2006. O quantificador *tudo* no PB. **Revista Letras**, n. 70, p. 157-182. Curitiba: Editora UFPR.

GODOY, M. C. 2019. **Introdução aos modelos lineares mistos para os estudos da linguagem**. PsyArXiv. Disponível em <https://doi.org/10.17605/OSF.IO/9T8UR> . Acesso em 09 fev. 2022.

GRIES, S. T. 2013. **Statistics for linguistics with R: A practical introduction**. Berlin: Walter de Gruyter.

GUERRA, S. et al. 2020. **Ciência de Dados com R – Introdução**. Bookdown: IBPAD. Disponível em: <https://cdr.ibpad.com.br/index.html> Acesso em 29 set. 2021.

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

GUY, G. 2000. A identidade linguística na comunidade de fala: paralelismo interdialetoal nos padrões de variação linguística. **Organon**, vol. 14 (28-29), 17-32, Disponível em <https://seer.ufrgs.br/index.php/organon/article/view/30194/18703>

GUY, G. 2005. A questão da crioulização no português do Brasil. In A. M. S. Zilles, (Ed.), **Estudos de Variação Lingüística no Brasil e no Cone Sul**, Editora da Universidade Federal do Rio Grande do Sul. pp. 15-38

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

HAUGEN, E. 1972. The Ecology of Language. In: DIL, A. S. (ed) **The Ecology of Language**: Essays by Einar Haugen, Stanford: Stanford University Press.

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

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censos demográficos**. 2010. Rio de Janeiro: IBGE. Disponível em: <http://biblioteca.ibge.gov.br/> Acesso em: 28 dez. 2021

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. 2018a. Cidades: Poções e Santo Antônio de Jesus. Disponível em: <http://www.ibge.gov.br/cidadesat/painel>. Acesso em 17 de mai 2019.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. 2018b. **Estatísticas de Gênero. Indicadores sociais das mulheres no Brasil**. Rio de Janeiro: IBGE. Disponível em: <https://www.ibge.gov.br/estatisticas/multidominio/genero/> Acesso em 28 dez. 2021.

KREUTZ, L. 2000. A educação de Imigrantes no Brasil. In: LOPES, E. M. T.; FARIA FILHO, L. M.; VEIGA, C. G. **500 anos de Educação no Brasil**. Belo Horizonte: Autêntica, p. 347-370.

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

LABOV, W. 1983. **Modelos Sociolingüísticos**. Madrid: ediciones Cátedra. Tradución de José Miguel Herreras.

LABOV, W. 2003. Some sociolinguistic principles. In: PAULSTON, C. B.; TUCKER, G. R. (Eds.). **Sociolinguistics: the essential readings**. Oxford: Blackwell, p. 235-250.

LEFEBVRE, C. 1997. Relexification in creole genesis: the case of demonstrative terms. In: Haitian creole. **Journal of pidgin and creole languages** 12:2, p.181 - 201.

LEFEBVRE, C. 1998. **Creole genesis and the acquisition of grammar: the case of Haitian Creole**. Cambridge: Cambridge University Press.

LEFEBVRE, C. 2001. Relexification in creole genesis and its effects on the development of the creole. **Journal of pidgin and creole languages**.

LEFEBVRE, C. 2008. On the principled nature of the respective contributions of substrate and superstrate languages to a creoles's lexicon. In: MICHAELIS, S. (ed.). **Roots of Creole Structures: Weighing the contribution of substrates and superstrates**, p. 197-223.

LEFEBVRE, C.; LUMSDEN, J. 1994. The central role of Relexification in Creole Genesis: The Case of Haitian Creole. In LEFEBVRE, Claire.; LUMSDEN, John (orgs.) **La genèse du créole haïtien: un cas particulier d'investigation sur la forma de la grammaire universelle**. UQAM, p. 59-180.

LEVSHINA, N. 2015. **How to do linguistics with R: Data exploration and statistical analysis**. Amsterdam: John Benjamins Publishing Company.

LIMA JR, R. M. *et al.* 2020. **Introdução a modelos de regressão para linguistas no R**. Disponível em: <https://guilhermegarcia.github.io/rling.html> Acesso em 27 dez 2021.

LIMA JR, R. M.; GARCIA, G. D. 2021. Diferentes análises estatísticas podem levar a conclusões categoricamente distintas. **Revista da Abralín**, v. 20, n.1, p. 1-19.

LEMLE, M.; NARO, A.1976. Syntatic diffusion. **Ciência e Cultura**, v. 29, n. 3, p.259-268.

LOPES, L. O. J. 2014. **A concordância nominal de número no português falado na zona rural de Santa Leopoldina**. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Linguística, Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://linguistica.ufes.br/pt-br/pos-graduacao/PPGEL/detalhes-da-tese?id=7770>. Acesso em 01 abril 2019.

LOPES, N. da S. 2001. **Concordância nominal, contexto lingüístico e sociedade**. Tese de doutorado. Universidade Federal da Bahia.

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. As Duas Grandes Vertentes da História Sociolinguística do Brasil. **DELTA**, São Paulo, v.17, n.1, 97-130.

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. 2015. **Língua e sociedade partidas: a polarização sociolinguística do Brasil**. São Paulo: Contexto.

LUCCHESI, D. 2019. Por que a criouliização aconteceu no Caribe e não no Brasil? Condicionamentos sócio-históricos. **Gragoatá**, Niterói, v.24, n. 48, p. 227-255.

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.

LUCCHESI, D.; DÁLIA, J. 2020. Equacionando o efeito da posição na variação da concordância nominal de número. **Revista Linguística**. UFRJ: Rio de Janeiro, v. 1, p. 771-798.

LUCCHESI, D.; DÁLIA, J. Novos condicionamentos estruturais da variação na concordância nominal de número. **Revista Fórum Linguístico**, Florianópolis, no prelo.

LUCCHESI, D.; MELLO, C. 2009. A alternância dativa. In: LUCCHESI, D.; BAXTER, A. N.; RIBEIRO, I. (orgs). **O português afro-brasileiro**. 427-456. Salvador: EDUFBA.



MATTOS E SILVA, R. V. 2004a. **Ensaio para uma sócio-história do português brasileiro**. São Paulo: Parábola Editorial.

MATTOS E SILVA, R. V. 2004b. **O português são dois**: novas fronteiras, velhos problemas. São Paulo: Parábola Editorial.

MATTOSO, K. 1982. **Ser escravo no Brasil**. São Paulo: Brasiliense.

McWHORTER, J. 2014. A response to Mufwene. **Journal of Pidgins and Creole Languages**. 29:1, p. 172-176.

MEISEL, J. M. 1983. Transfer as a second-language strategy. **Language & Communication**, 3(1), p. 11–46.

MELATTI, J. C. 2014. **Índios do Brasil**. São Paulo: Edusp, 2014.

MENDES, G. F; ALMEIDA, M. G. de. 2008. Memória, símbolos e representações na configuração socioespacial do Sertão da Ressaca – Bahia. **Revista Mercator**, Vol. 7, n.13, p. 29-37.

MENDES, R. B. 2014. Gendered perceptions of noun agreement in Brazilian Portuguese. **Revista Interacional de Linguística Iberoamericana**, vol. 12(23), 93-108.

MENÉNDEZ, Miguel. 1982. Uma contribuição para a etno-história da árca Tapajós-Madeira. **Revista do Museu Paulista**, 1.s., p. 28-289.

MICHAELIS, S. *et al.* 2013. **The Atlas of Pidgin and Creole Structures**. Oxford: Oxford University Press.

MONDLANE, J. A. 2012. **Sistemas de contagem em Changana: alguns aspectos morfossintáticos e semânticos dos numerais**. Dissertação (Mestrado em Linguística). Faculdade de Letras e Ciências Sociais. Universidade Eduardo Mondlane, Maputo.

MONTEIRO, N. G. 2010. Idade Moderna (séculos XV-XVIII). In: RAMOS, R. *et. al.* (org.). **História de Portugal**. Lisboa: A esfera dos Livros, p. 197-435.

MUFWENE, S. S. 2001. **The ecology of language evolution**. Cambridge: Cambridge University Press.

MUFWENE, S. S. 2002. Competition and Selection in Language Evolution. In **Selections**. Chicago, p. 45 - 56. Disponível em: <http://www.akkr.hu/journals/select> Acesso em 07 de mai. 2021.

MUFWENE, S. S. 2008. **Language Evolution**: contact, competition and change. London: Continuum.

MUFWENE, S. S. 2018. Language Evolution from an Ecological Perspective. **The Routledge Handbook of Ecolinguistics**. Ed. by Alwin F. Fill & Hermine Penz. p. 73-88. Disponível em: <https://www.researchgate.net> Acesso em 05 de fev. 2021

MÜLLER, A. 2002. Nomes nus e o parâmetro nominal. **Revista Letras**. Curitiba, nº 58, p. 325 – 337. Editora UFPR. Disponível em <https://revistas.ufpr.br/letras> Acesso em 12 fev. 2022.

MUSSA, A. 1991. **O papel das línguas africanas na história do português do Brasil**. Dissertação (Mestrado em Letras) Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2 v.

MUYSKEN, P. 2001. The origin of creole languages: The perspective of second language learning. In **Creolization and Contact**. Norval Smith and Tonjes Veenstra (eds). Amsterdam: John Benjamins. 157–173.

NARO, A. 1981. The social and structural dimensions of syntactic change. **Language**, v. 57, n.1, p. 63-98.

NARO, A.; SCHERRE, M. M.P. 1991. Variação e mudança linguística: fluxos e contrafluxos na comunidade de fala. **Caderno de Estudos Linguísticos, Campinas**, vol. 20 (jan/jun), 9-16 Disponível em <https://periodicos.sbu.unicamp.br/ojs/index.php/cel>

NARO, A.; SCHERRE, M. M. P. 1993. Sobre as origens do português popular do Brasil. **DELTA**, n. 9, 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.

NEGRÃO, E. V.; VIOTTI, E. 2012. Em busca de uma História Linguística. **Revista de Estudos da Linguagem**. Belo Horizonte, v. 20, n.2, p. 309-42.

NGUNGA, A. 2004. **Introdução à Linguística Bantu**. Imprensa Universitária: Faculdade de Letras e Ciências Sociais.

NUZZO, R. 2014 Scientific method: statistical errors. **Nature News**, v. 506, n. 7487, p. 150.

OUSHIRO, L. 2015. Identidade na pluralidade: avaliação, produção e percepção linguística na cidade de São Paulo. Tese (Doutorado em Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.

- OUSHIRO, L. 2017. **Introdução à Estatística para Linguistas**, v.1.0.1 (dez/2017). Disponível em: <https://rstudio-pubs-static.s3.amazonaws.com/> Licença Creative Commons 4.0 Atribuição – Não comercial. Acesso em 15 jan. 2022
- OUSHIRO, L. 2021. **Introdução à Estatística para Linguistas** (Version 2.0.3). Zenodo. Disponível em <http://doi.org/10.5281/zenodo.4755739>. 2021. Acesso em 16 jan 2022.
- PAWLAK, N. 2012. Morphological Coding of Verb-Object Agreement in African Languages. **Rocznik Orientalistyczny**, T. LXV, Z. 1, p. 158–170.
- PESSOA DE CASTRO, Y. 2006. A matriz africana no português do Brasil. In: CARDOSO, S.; MOTA, J.; MATTOS E SILVA, R. V. (Org.). **Quinhentos Anos de História Lingüística do Brasil**. Salvador: Secretaria da Cultura e Turismo do Estado da Bahia, p. 81-116.
- PETTER, M. M. T. 2006. Línguas Africanas no Brasil. In: CARDOSO, S.; MOTA, J.; MATTOS E SILVA, R. V. (Org.). **Quinhentos Anos de História Lingüística do Brasil**. Salvador: Secretaria da Cultura e Turismo do Estado da Bahia, p. 17-142.
- PIENEMANN, M. et al. 2005 Processability, typological distance and L1 transfer. In: **Cross-Linguistic Aspects of Processability Theory**. Manfred Pienemann (ed.). Studies in Bilingualism, 30. 85–116.
- PLAG, I. 2008. Creoles as interlanguages: Syntactic structures. **Journal of Pidgin and Creole Languages**. 23: 2, 307–238.
- PROJETO VERTENTES. Disponível em: <<http://www.vertentes.ufba.br/>>. Acesso em: 23 mar. 2019.
- R Core Team. 2021. **R: a language and environment for statistical computing**. Viena: R Foundation for Statistical Computing. Disponível em [https:// www.R-project.org](https://www.R-project.org). Acesso em 01 de nov. 2021.
- RADFORD, A. 1997. **Syntactic Theory and the Structure of English**. Cambridge: Cambridge University Press.
- REIS, J. J. 2007. Presença negra: conflitos e encontros. In: **Brasil: 500 anos de povoamento**. IBGE, Centro de Documentação e Disseminação de Informações. - Rio de Janeiro: IBGE, p. 79-100.
- RIBEIRO, C. M. da. 2018. **Contato Lingüístico e a Concordância de Número no Sintagma Nominal no Português de Oiapoque-AP**. Tese de doutorado. Universidade Federal do Rio de Janeiro.
- RODRIGUES, A. D. 1993. Línguas indígenas: 500 anos de descobertas e perdas. **DELTA**., vol. 9, n. 1, p. 83-103.
- RODRIGUES, A. D. 1996. As línguas gerais sul-americanas. **PAPIA** - Revista de crioulos de base ibérica, v. 4, p. 6-18.

RODRIGUES, A. D. 2000. Panorama das Línguas Indígenas da Amazônia. In: F. Queixalós & Renault-Lescure (orgs.). **As línguas amazônicas hoje**. São Paulo: IRD/ISA/MPEG, p. 15-28.

RODRIGUES, A. D. 2005. Sobre as línguas indígenas e sua pesquisa no Brasil. **Ciência e Cultura**. vol.57, n.2, p. 35-38.

SALOMÃO, M. H. 2010. **A variação de pluralidade nas estruturas predicativas da variedade falada na região de São José do Rio Preto**. Dissertação de mestrado. São José do Rio preto: UNESP, IBILCE. 162 f. Disponível em <http://base.repositorio.unesp.br/handle/11449/86554>

SANTOS, L. L. 2016. **Fala (você/tu) ~ fale (você/tu): a expressão variável do modo imperativo no português popular da Bahia**. Tese de doutorado. Universidade Federal da Bahia.

SAUSSURE, F. 1995. **Curso de lingüística geral**. São Paulo: Editora Cultrix LTDA, [1916]. (Tradução de Antônio Chelini, José Paulo Paes e Izidoro Blikstein).

SCHEERE, M. M. P. 1978. **A regra de concordância de número no sintagma nominal em português**. Dissertação de mestrado. Rio de Janeiro: Faculdade de Letras da UFRJ. 158f.

SCHERRE, M. M. P. 1988. **Reanálise da concordância nominal em português**. Tese de Doutorado. Universidade Federal do Rio de Janeiro, Rio de Janeiro.

SCHERRE, M. M. P. 1997. Concordância nominal e funcionalismo. **Alfa**. São Paulo, 41 (n. esp.) 181-206.

SCHÜTZ, R. E. 2018. **Interlíngua e Fossilização**. English Made in Brazil . Acesso em 04 dez. 2019.

SCHWARCZ, L. M.; STARLING, H. M. 2015. **Brasil: uma biografia**. São Paulo: Companhia das Letras.

SILVA, J. A. A. 2005. **A concordância verbal de terceira pessoa do plural no português popular do Brasil: um panorama sociolingüístico de três comunidades do interior do estado da Bahia**. 2005. 324 f. Tese de doutorado. Universidade Federal da Bahia.

SILVA, J. B. da; SCHERRE, M. M. P. 2013. A concordância nominal na fala capixapa: fatores sociais. In: CARDOSO, C. R.; SCHERRE, M. M.; LIMA-SALLES, H. M.M.; PACHECO, C. (orgs). **Variação linguística: contato de línguas e educação**, p. 129-143. Campinas: Pontes Editores.

SILVA, J. A. A. da; SOUSA, V. V. 2013. Pelo “Sertão da Ressaca”: contribuições para a compreensão da sócio-história do Português Popular do Brasil. **Revista Tabuleiro de Letras** – UNEB. Departamento de Ciências Humanas – DCH I, n. 6. Disponível em: <<https://www.revistas.uneb.br/>> Acesso em 06 set. 2021.

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

TARALLO, F. 1993. Sobre a alegada origem crioula do português brasileiro: mudanças sintáticas aleatórias. In: ROBERTS, Ian; KATO, Mary (orgs). **Português Brasileiro**: uma viagem diacrônica. 2. ed. Campinas: Editora da UNICAMP (versão original em inglês 1986). p.35-68.

THOMASON, S.G., 2001. **Language Contact**: An Introduction. Georgetown University Press, Washington, DC.

THOMASON, S.G.; KAUFMAN, T., 1988. Language Contact, Creolization, and Genetic Linguistics. **Journal of Pidgin and Creole Languages**, 5(1).

TRANNIN, J. B. 2016. **Tudo na história do português**. Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas, SP. Disponível em: <http://www.repositorio.unicamp.br/handle/REPOSIP/305670> Acesso em: 26 dez. 2021.

TRANNIN, J. B. 2017. As duas gramáticas de *tudo* e as variedades do português afro-brasileiro. **Revista Letras Escreve**, Macapá, v.7, n.2, p. 135-160.

TRASK, R. L. 2004. **Dicionário de linguagem e linguística**. Trad. Rodolfo Ilari. São paulo: Contexto.

VIEIRA, S. R.; BRANDÃO, S. F. 2014. Tipologia de Regras Linguísticas e Estatuto das Variedades/Línguas: a Concordância em Português. **Linguística**, v. 30 (2), 81-112.

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.

WINFORD, D. 2003. **An Introduction to contact linguistics**. Oxford: Blackwell. Disponível em: [www.R-project.org](http://www.R-project.org). Acesso em 01 nov. 2021.

XATARA, C. M. ; OLIVEIRA, W. L. de. 2002. **Dicionário de provérbios, idiomatismos e palavrões**: francês-português / português-francês. São Paulo: Cultura.

YACOVENCO, L. C. *et all.* 2012. Projeo Portvix: A fala de Vitória/ES em cena. **Alfa**, vol 56(3), 771-806.

## SCRIPT DO R - Concordância nominal no português do interior da Bahia

```

setwd("C:/Users/Thamíris/Desktop/MEUS DADOS")
library(tidyverse)
library(ggplot2)
library(effects)
library(lme4)
library(car)
dados1 <- read.csv2("DADOS.GERAIS final.csv",
  stringsAsFactors = TRUE,
  header = T, dec = ",")
glimpse(dados1)
summary(dados1)
View(dados1)
str(dados1)
tab.vd <- table(dados1$OCORRENCIA)
tab.vd

#VARIÁVEL DEPENDENTE CN (AUSÊNCIA X PRESENÇA)#

tab.prop.vd <- dados1 %>%
  count(OCORRENCIA) %>%
  mutate(prop = prop.table(n)*100) %>%
  print()

ggplot(tab.prop.vd, aes(x = OCORRENCIA, y = prop, fill = OCORRENCIA)) +
  geom_bar(stat = "identity", color = "blue") +
  ggtitle("OCORRÊNCIAS") +
  labs(x = "Regra variável da CN", y = "Proporção", fill = "Regra variável da CN") +
  scale_x_discrete(labels = c("ausência", "presença")) +
  scale_fill_brewer(palette = "Purples", labels = c("presença", "ausência")) +
  theme_bw()
dados <- read.csv2("DADOS0902.csv",

```

```
stringsAsFactors = TRUE,
header = T, dec = ",")
```

```
saveRDS(dados, file = "dadosthamiris.Rds")
```

```
tab.prop.vd <- dados %>%
  count(OCORRENCIA) %>%
  mutate(prop = prop.table(n)*100) %>%
  print()
```

```
ggplot(tab.prop.vd, aes(x = OCORRENCIA, y = prop, fill = OCORRENCIA)) +
  geom_bar(stat = "identity", color = "blue") +
  ggtitle("OCORRÊNCIAS") +
  labs(x = "Regra variável da CN", y = "Proporção", fill = "Regra variável da CN") +
  scale_x_discrete(labels = c("ausência", "presença")) +
  scale_fill_brewer(palette = "Purples", labels = c("presença", "ausência")) +
  theme_bw()
```

```
#VARIÁVEIS LINGUÍSTICAS####
```

```
#POSIÇÃO
```

```
tab.posicao <- table(dados$POSICAO.DA.MARCA, dados$OCORRENCIA)
tab.posicao
```

```
posicao <- dados %>%
  group_by(OCORRENCIA, POSICAO.DA.MARCA) %>%
  summarise(posicao=n()) %>%
  filter(OCORRENCIA == "ausencia") %>%
  mutate(perc=posicao/sum(posicao)*100)
posicao
```

```
#CLASSE DA POSIÇÃO MARCADA
```

```
tab.classe <- table(dados$CLASSE.POSICAO, dados$OCORRENCIA)
```

```
tab.classe
```

```
classe <- dados %>%
```

```
  group_by(OCORRENCIA, CLASSE.POSICAO) %>%
```

```
  summarise(classe=n()) %>%
```

```
  filter(OCORRENCIA == "ausencia") %>%
```

```
  mutate(perc=classe/sum(classe)*100)
```

```
classe
```

```
view(classe)
```

```
##CONFIGURAÇÃO SINTAGMÁTICA DO SN#
```

```
prop.config <- dados %>%
```

```
  group_by(OCORRENCIA, CONFIGURACAO.SN) %>%
```

```
  summarise(prop.config=n()) %>%
```

```
  filter(OCORRENCIA == "presenca") %>%
```

```
  mutate(prop = prop.config/sum(prop.config)*100)
```

```
prop.config
```

```
tab.prop.config <- dados %>%
```

```
  count(CONFIGURACAO.SN,OCORRENCIA) %>%
```

```
  group_by(CONFIGURACAO.SN) %>%
```

```
  mutate(prop = prop.table(n)*100) %>%
```

```
  print()
```

```
View(tab.prop.config)
```

```
levels(dados$CONFIGURACAO.SN)
```

```
ggplot(tab.prop.config, aes(x = CONFIGURACAO.SN, y = prop, fill = OCORRENCIA))
```

```
+
```

```
  geom_bar(stat = "identity", color = "black") +
```



```

ggtitle("CONFIGURAÇÃO SINTAGMÁTICA DO SN") +
labs(x = "CONFIGURAÇÃO SINTAGMÁTICA DO SN", y = "Proporção", fill = "Regra
variável da CN") +
scale_x_discrete(labels =
c("adjetivo.nome", "det.nome", "nome.adjetivo", "numer.nome", "quant.nome")) +
scale_fill_brewer(palette = "Purples", labels = c("ausência", "presença")) +
theme_bw()

```

```

tab.config <- with(dados, table(CONFIGURACAO.SN, OCORRENCIA))

```

```

tab.config

```

```

q.testeconfig <- chisq.test(tab.config)

```

```

q.testeconfig

```

## ##REALIZAÇÃO DO NÚCLEO

```

tab.prop.nucleo <- dados %>%
count(NUCLEO,OCORRENCIA) %>%
group_by(NUCLEO) %>%
mutate(prop = prop.table(n)*100) %>%
print()

```

```

ggplot(tab.prop.nucleo, aes(x = NUCLEO, y = prop, fill = OCORRENCIA)) +
geom_bar(stat = "identity", color = "black") +
ggtitle("REALIZAÇÃO DO NÚCLEO") +
labs(x = "REALIZAÇÃO DO NÚCLEO", y = "Proporção", fill = "Regra variável da CN")
+
scale_x_discrete(labels = c("com.nucleo", "sem.nucleo")) +
scale_fill_brewer(palette = "Purples", labels = c("ausência", "presença")) +
theme_bw()

```

```

tab.nucleo <- with(dados, table(NUCLEO, OCORRENCIA))

```

```

tab.nucleo

```

```

q.testenucleo <- chisq.test(tab.nucleo)

```

```

q.testenucleo

```

### ##NÚMERO DE CONSTITUINTES DO SN##

```
tab.prop.constituientes <- dados %>%
  count(NUMERO.CONSTITUINTES,OCORRENCIA) %>%
  group_by(NUMERO.CONSTITUINTES) %>%
  mutate(prop = prop.table(n)*100) %>%
  print()
```

```
levels(dados$NUMERO.CONSTITUINTES)
```

```
ggplot(tab.prop.constituientes, aes(x = NUMERO.CONSTITUINTES, y = prop, fill =
OCORRENCIA)) +
  geom_bar(stat = "identity", color = "purple") +
  ggtitle("NÚMERO DE CONSTITUINTES NO SN") +
  labs(x = "NÚMERO DE CONSTITUINTES NO SN", y = "Proporção", fill = "Regra
variável da CN") +
  scale_x_discrete(labels = c("dois", "quatro ou mais", "três")) +
  scale_fill_brewer(palette = "Purples", labels = c("ausência", "presença")) +
  theme_bw()
```

```
tab.constituientes <- with(dados, table(NUMERO.CONSTITUINTES, OCORRENCIA))
tab.constituientes
q.testeconstituientes <- chisq.test(tab.constituientes)
q.testeconstituientes
```

### ##NÚMERO DE CONSTITUINTES FLEXIONÁVEIS DO SN##

```
tab.prop.flex <- dados %>%
  count(NUMERO.FLEXIONAVEIS,OCORRENCIA) %>%
  group_by(NUMERO.FLEXIONAVEIS) %>%
  mutate(prop = prop.table(n)*100) %>%
  print()
```

```
levels(dados$NUMERO.FLEXIONAVEIS)
```

```
ggplot(tab.prop.flex, aes(x = NUMERO.FLEXIONAVEIS, y = prop, fill =
OCORRENCIA)) +
  geom_bar(stat = "identity", color = "black") +
  ggtitle("NÚMERO DE CONSTITUINTES FLEXIONÁVEIS") +
  labs(x = "NÚMERO DE CONSTITUINTES FLEXIONÁVEIS", y = "Proporção", fill =
"Regra variável da CN") +
  scale_x_discrete(labels = c("dois.flex","quatro.mais.flex","tres.flex", "um.flex")) +
  scale_fill_brewer(palette = "Purples", labels = c("ausência", "presença")) +
  theme_bw()
```

```
tab.flex <- with(dados, table(NUMERO.FLEXIONAVEIS, OCORRENCIA))
```

```
tab.flex
```

```
q.testeflex <- chisq.test(tab.flex)
```

```
q.testeflex
```

```
##SALIÊNCIA FÔNICA
```

```
tab.prop.saliencia <- dados %>%
  count(SALIENCIA,OCORRENCIA) %>%
  group_by(SALIENCIA) %>%
  mutate(prop = prop.table(n)*100) %>%
  print()
```

```
levels(dados$SALIENCIA)
```

```
ggplot(tab.prop.saliencia, aes(x = SALIENCIA, y = prop, fill = OCORRENCIA)) +
  geom_bar(stat = "identity", color = "black") +
  ggtitle("SALIÊNCIA FÔNICA") +
  labs(x = "SALIÊNCIA FÔNICA", y = "Proporção", fill = "Regra variável da CN") +
  scale_x_discrete(labels = c("com.saliencia", "sem.saliencia")) +
  scale_fill_brewer(palette = "Purples", labels = c("ausência", "presença")) +
  theme_bw()
```

```
tab.saliencia <- with(dados, table(SALIENCIA, OCORRENCIA))
```

```
tab.saliencia
```

```
q.testesaliencia <- chisq.test(tab.saliencia)
```

```
q.testesaliencia
```

```
##VARIÁVEIS SOCIAIS##
```

```
#SEXO#
```

```
tab.prop.sexo <- dados %>%
```

```
  count(SEXO,OCORRENCIA) %>%
```

```
  group_by(SEXO) %>%
```

```
  mutate(prop = prop.table(n)*100) %>%
```

```
  print()
```

```
ggplot(tab.prop.sexo, aes(x = SEXO, y = prop, fill = OCORRENCIA)) +
```

```
  geom_bar(stat = "identity", color = "black") +
```

```
  ggtitle("SEXO") +
```

```
  labs(x = "SEXO", y = "Proporção", fill = "Regra variável da CN") +
```

```
  scale_x_discrete(labels = c("feminino", "masculino")) +
```

```
  scale_fill_brewer(palette = "Purples", labels = c("ausência", "presença")) +
```

```
  theme_bw()
```

```
tab.sexo <- with(dados, table(SEXO, OCORRENCIA))
```

```
tab.sexo
```

```
q.testesexo <- chisq.test(tab.sexo)
```

```
q.testesexo
```

```
##FAIXA ETÁRIA##
```

```
tab.prop.faixaetaria <- dados %>%
```

```
  count(FAIXA.ETARIA,OCORRENCIA) %>%
```

```
group_by(FAIXA.ETARIA) %>%
mutate(prop = prop.table(n)*100) %>%
print()
```

```
levels(dados$FAIXA.ETARIA)
```

```
ggplot(tab.prop.faixaetaria, aes(x = FAIXA.ETARIA, y = prop, fill = OCORRENCIA)) +
  geom_bar(stat = "identity", color = "black") +
  ggtitle("FAIXA ETÁRIA") +
  labs(x = "FAIXA ETÁRIA", y = "Proporção", fill = "Regra variável da CN") +
  scale_x_discrete(labels = c("idoso", "intermediária", "jovem")) +
  scale_fill_brewer(palette = "Purples", labels = c("ausência", "presença")) +
  theme_bw()
```

```
tab.faixaetaria <- with(dados, table(FAIXA.ETARIA, OCORRENCIA))
```

```
tab.faixaetaria
```

```
q.testefaixaetaria <- chisq.test(tab.faixaetaria)
```

```
q.testefaixaetaria
```

```
##ESCOLARIDADE##
```

```
tab.prop.escolaridade <- dados %>%
  count(ESCOLARIDADE,OCORRENCIA) %>%
  group_by(ESCOLARIDADE) %>%
  mutate(prop = prop.table(n)*100) %>%
  print()
```

```
ggplot(tab.prop.escolaridade, aes(x = ESCOLARIDADE, y = prop, fill =
OCORRENCIA)) +
  geom_bar(stat = "identity", color = "black") +
  ggtitle("ESCOLARIDADE") +
  labs(x = "ESCOLARIDADE", y = "Proporção", fill = "Regra variável da CN") +
  scale_x_discrete(labels = c("pouca.escolaridade", "sem.escolaridade")) +
```

```
scale_fill_brewer(palette = "Purples", labels = c("ausência", "presença")) +
theme_bw()
```

```
levels(dados$ESCOLARIDADE)
```

```
tab.escolaridade <- with(dados, table(ESCOLARIDADE, OCORRENCIA))
```

```
tab.escolaridade
```

```
q.testeescolaridade <- chisq.test(tab.escolaridade)
```

```
q.testeescolaridade
```

```
##ESTADIA FORA DA COMUNIDADE##
```

```
tab.prop.estadia <- dados %>%
```

```
  count(ESTADIA,OCORRENCIA) %>%
```

```
  group_by(ESTADIA) %>%
```

```
  mutate(prop = prop.table(n)*100) %>%
```

```
  print()
```

```
levels(dados$ESTADIA)
```

```
ggplot(tab.prop.estadia, aes(x = ESTADIA, y = prop, fill = OCORRENCIA)) +
```

```
  geom_bar(stat = "identity", color = "black") +
```

```
  ggtitle("ESTADIA") +
```

```
  labs(x = "ESTADIA", y = "Proporção", fill = "Regra variável da CN") +
```

```
  scale_x_discrete(labels = c("nao.viveu.fora", "viveu.fora")) +
```

```
  scale_fill_brewer(palette = "Purples", labels = c("ausência", "presença")) +
```

```
  theme_bw()
```

```
tab.estadia <- with(dados, table(ESTADIA, OCORRENCIA))
```

```
tab.estadia
```

```
q.testeestadia <- chisq.test(tab.estadia)
```

```
q.testeestadia
```

```
##LOCALIDADE##
```

```

tab.prop.localidade <- dados1 %>%
  count(LOCALIDADE,OCORRENCIA) %>%
  group_by(LOCALIDADE) %>%
  mutate(prop = prop.table(n)*100) %>%
  print()

```

```

levels(dados1$LOCALIDADE)

```

```

ggplot(tab.prop.localidade, aes(x = LOCALIDADE, y = prop, fill = OCORRENCIA)) +
  geom_bar(stat = "identity", color = "purple") +
  ggtitle("LOCALIDADE") +
  labs(x = "LOCALIDADE", y = "Proporção", fill = "Regra variável da CN") +
  scale_x_discrete(labels
  =
c("rural.pocoas","rural.santoant","sede.pocoas","sede.santoant")) +
  scale_fill_brewer(palette = "Purples", labels = c("ausência", "presença")) +
  theme_bw()

```

```

tab.prop.localidade <- dados %>%
  count(LOCALIDADE,OCORRENCIA) %>%
  group_by(LOCALIDADE) %>%
  mutate(prop = prop.table(n)*100) %>%
  print()

```

```

levels(dados$LOCALIDADE)

```

```

tab.localidade <- with(dados, table(LOCALIDADE, OCORRENCIA))
tab.localidade
q.testelocalidade <- chisq.test(tab.localidade)
q.testelocalidade

```

```
##INFORMANTE##
```

```
tab.prop.informante <- dados %>%  
  count(INFORMANTE,OCORRENCIA) %>%  
  group_by(INFORMANTE) %>%  
  mutate(prop = prop.table(n)*100) %>%  
  print()
```

```
#REGRESSÃO LOGÍSTICA
```

```
mod5 <- glmer(OCORRENCIA ~ CONFIGURACAO.SN +  
  NUCLEO +  
  SALIENCIA +  
  SEXO +  
  FAIXA.ETARIA +  
  ESCOLARIDADE +  
  LOCALIDADE +  
  ESTADIA +  
  (1|INFORMANTE),  
  family = binomial,  
  data = dados)  
summary(mod5)  
car::vif(mod5)  
plot(allEffects(mod5), ask = T, type = "response")
```

```
mod6 <- glmer(OCORRENCIA ~ SALIENCIA +  
  NUCLEO +  
  NUMERO.FLEXIONAVEIS +  
  SEXO +  
  FAIXA.ETARIA +  
  ESCOLARIDADE +  
  LOCALIDADE +  
  ESTADIA +  
  (1|INFORMANTE),
```



```
        family = binomial,
        data = dados)
summary(mod6)
car::vif(mod6)
plot(allEffects(mod6), ask = T, type = "response")
drop1(mod6, teste = "LR")

tab.cruzamento <- table(dados$POSICAO.DA.MARCA, dados$CLASSE.POSICAO,
dados$OCORRENCIA)
tab.cruzamento
tab.prop.cruzamento <- prop.table(tab.cruzamento, 1)*100
tab.prop.cruzamento

tab.det_nucleo <- table(dados$NUCLEO, dados$PRESENCA.DET,
dados$OCORRENCIA)
tab.det_nucleo
tab.prop.det_nucleo <- prop.table(tab.det_nucleo, 1)*100
tab.prop.det_nucleo

tab.numero <- table(dados$NUMERO.CONSTITUINTES,
dados$NUMERO.FLEXIONAVEIS, dados$OCORRENCIA)
tab.numero
tab.prop.numero <- prop.table(tab.numero, 1)*100
tab.prop.numero

tab.escola_faixa <- table (dados$ESCOLARIDADE, dados$FAIXA.ETARIA,
dados$OCORRENCIA)
tab.escola_faixa
tab.prop.escola_faixa <- prop.table(tab.escola_faixa, 1)*100
tab.prop.escola_faixa

tab.saliencia_faixa <- table(dados$SALIENCIA, dados$FAIXA.ETARIA,
dados$OCORRENCIA)
tab.saliencia_faixa
```

```
tab.prop.saliencia_faixa <- prop.table(tab.saliencia_faixa)
```

```
tab.prop.saliencia_faixa
```

```
tab.saliencia_escola <- table(dados$SALIEDNCIA, dados$ESCOLARIDADE,  
dados$OCORRENCIA)
```

```
tab.saliencia_escola
```

```
tab.prop.saliencia_escola <- prop.table(tab.saliencia_escola)
```

```
tab.prop.saliencia_escola
```