



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE ARTES E COMUNICAÇÃO
DEPARTAMENTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

THAYSE CAROLINA FERREIRA PARAISO

**O USO DA CONCORDÂNCIA NOMINAL DE NÚMERO EM REDAÇÕES
PRODUZIDAS EM PORTUGUÊS POR ESTUDANTES GUINEENSES E
TIMORENSES: um estudo contrastivo**

Recife
2019

THAYSE CAROLINA FERREIRA PARAISO

**O USO DA CONCORDÂNCIA NOMINAL DE NÚMERO EM REDAÇÕES
PRODUZIDAS EM PORTUGUÊS POR ESTUDANTES GUINEENSES E
TIMORENSES: um estudo contrastivo**

Dissertação apresentada ao Programa de Pós-Graduação em Letras da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestra em Letras.

Área de concentração: Linguística

Orientadora: Prof^ª. Dr^ª. Cláudia Roberta Tavares Silva

Recife

2019

Catálogo na fonte
Bibliotecária Jéssica Pereira de Oliveira, CRB-4/2223

P222u Paraíso, Thayse Carolina Ferreira
O uso da concordância nominal de número em redações produzidas em português por estudantes guineenses e timorenses: um estudo contrastivo / Thayse Carolina Ferreira Paraíso. – Recife, 2019. 126f.: il.

Orientadora: Cláudia Roberta Tavares Silva.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. Centro de Artes e Comunicação. Programa de Pós-Graduação em Letras, 2019.

Inclui referências.

1. Concordância nominal. 2. Variação. 3. Variedades do português.
I. Silva, Cláudia Roberta Tavares (Orientadora). II. Título.

410 CDD (22. ed.) UFPE (CAC 2020-41)

THAYSE CAROLINA FERREIRA PARAISO

**O USO DA CONCORDÂNCIA NOMINAL DE NÚMERO EM REDAÇÕES
PRODUZIDAS EM PORTUGUÊS POR ESTUDANTES GUINEENSES E
TIMORENSES: um estudo contrastivo**

Dissertação apresentada ao Programa de Pós-Graduação em Letras da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestra em Letras.

Aprovada em: 19/02/2019.

BANCA EXAMINADORA

Prof^a. Dr^a. Cláudia Roberta Tavares Silva (Orientadora)
Universidade Federal de Pernambuco

Prof. Dr. Marcelo Amorim Sibaldo (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Adeilson Pinheiro Sedrins (Examinador Externo)
Universidade Federal Rural de Pernambuco

AGRADECIMENTOS

A Deus, força motriz, pelo dom da vida, que me permite exercer, com satisfação, a profissão que escolhi e continuar a minha formação acadêmica; e a Nossa Senhora, minha mãe Maria, pelo discernimento e intercessão concedidos mesmo quando eu não sabia que queria ou que precisava.

A melhor mãe que Deus poderia ter me dado a honra de nascer, Marluce Ferreira, minha mãinha, por tudo que fez e faz por mim nessa vida, passando por cima, diversas vezes, de suas vontades para atender as minhas. Por todo o mimo que ainda recebo, como as comidinhas gostosas e cheias de afeto que sempre me tiram um pouco da frente do computador. Pela compreensão com as minhas ausências e falta de tempo. A senhora é a melhor mãe que eu poderia ter e eu sou muito, muito grata a Deus por isso.

A Denys Paraiso, painho, pelo amor incondicional que me dedica, e por ter me ensinado que o conhecimento é a melhor herança que se pode oferecer a um filho.

À minha querida professora e orientadora, Cláudia Roberta Tavares Silva, pelo trabalho espetacular de orientação e por toda a atenção, PACIÊNCIA, compreensão, torcida e ENCORAJAMENTO oferecidos durante as aulas e durante o período de realização da pesquisa que resultou nesta dissertação. Nunca esquecerei o desejo de “muito sucesso sempre”, presente na despedida de todos os diversos e-mails trocados. Eu também não poderia ter orientadora melhor.

Ao professor Marcelo Amorim Sibaldo, por ter compartilhado seus saberes durante as aulas e reuniões do GETEGRA, e por ter aceitado o convite para participar da banca avaliadora na defesa.

Ao professor Adeilson Pinheiro Sedrins, que acompanhou e contribuiu ricamente com este trabalho desde a fase de qualificação, e também aceitou o convite para participar da banca avaliadora na defesa.

Ao professor Vicente Massip, por ter compartilhado seus conhecimentos durante as aulas e por ter aceitado fazer parte da banca avaliadora na defesa, como membro suplente. Agradeço, também, à professora Dorothy Bezerra, por ter aceitado participar, como membro suplente, da banca avaliadora deste trabalho.

Às queridas amigas que o mestrado me deu de presente para a vida, Elizabeth Costa e Gabryella Fraga, pela parceria em TODOS os momentos durante

esses dois anos. A frase: “sem vocês isso não teria sido possível” nunca fez tanto sentido. Que bom que temos a vida inteira para curtirmos a nossa amizade.

Aos demais companheiros de turma, em especial os amigos Daniela, Jamilys, Edrielly e Juliana, pela parceria durante as aulas e por todas as palavras e atitudes de apoio e incentivo.

Aos amigos Dereck e Alane, por dividirem conosco o que sabiam sobre o programa *GoldVarb-X* e por estarem sempre disponíveis para ajudar com os estudos em Sociolinguística. E falando sobre estudos em Sociolinguística, e sobre torcida, apoio e incentivo, agradeço, também, aos amigos Cícero e Flávia, irmãos por parte de orientadora.

À UNILAB-CE, que agradeço na pessoa da pró-reitora de graduação, a professora Andréa Linard, por ter dado anuência para que eu pudesse coletar as redações de vestibular que analisei nesta pesquisa, e por ter me recebido tão bem ao chegar lá, além de me deixar nas mãos da melhor equipe, alguns dos servidores da PROGRAD: Isabelle, Mayara e Dante; e as funcionárias Margarida (que ajudou MUITO no acesso ao banco de textos) e Jéssica (que me apresentou a cidade). Aproveito para agradecer, também, ao ex-presidente Lula, por ter implementado esta Universidade no Brasil. Lula livre!

A Vera, Luciana, Thainan e Thuany, minhas primas queridas, pelo carinho, apoio, incentivo e torcida de sempre, além do orgulho (sem razão) que sentem de mim. Eu que me orgulho em ter vocês na vida. Obrigada pelo amor de FAMÍLIA.

Às amigas Jéssica e Diva, grandes amores dessa vida, pelas portas sempre abertas da casa, do colo e do coração.

Aos amigos Alex Inácio e Andréa Barretto, pela ajuda no final do processo de escrita, fase crucial em que nos falta o ar. Vocês foram combustível extra.

Aos amigos do IFPE, Douglas, Josemar e Maria Clara, que me ajudaram MUITO na seleção para ingressar no mestrado, e seguiram ajudando durante a realização da pesquisa.

Aos demais queridos colegas professores de português do NLP-IFPE, grandes companheiros de trabalho, que permitiram o meu afastamento parcial das atividades profissionais, assumindo parte das minhas atividades e deixando essa fase de estudo mais leve. Agradeço, especialmente, ao meu amigo e coordenador de área, o professor Adriano Moura, e repito o agradecimento à minha amiga e parceira de VÁRIOS trabalhos, a professora Maria Clara Catanho, pelas conversas

sobre a minha pesquisa e pela pré-banca que fizeram comigo, ajudando-me a conseguir aprovação, inclusive, no doutorado deste programa.

Aos amigos de longa data, Ana Evelyn, Ana Luísa (também minha comadre querida), Ayla e Michelli, por estarem sempre por perto e por compreenderem minhas ausências no período de reclusão e solidão que foi essa escrita.

Às amigas queridas, Xênia, Andreza e Thamiris, que contribuíram com seu carinho e sua torcida para que tudo desse certo e não deixaram de perguntar da dissertação, lembrando-me que eu estava procrastinando e que deveria ter foco e disciplina.

A todos os professores do PPGL-UFPE, sobretudo aqueles que ministraram as disciplinas que cursei, Medianeira, Virgínia, Sibaldo, Cláudia, Xavier, Joice, Kasuê e Massip, pelo bom exemplo de profissionalismo e pelos ensinamentos compartilhados durante esse período de muito aprendizado.

E, finalmente, um agradecimento mais que especial ao pessoal da secretaria da pós-graduação em Letras, em especial a Jozaías e a Claudivane, sempre solícitos para resolver nossos “perrengues” e também grandes incentivadores nessa jornada.

A todos vocês, MUITO OBRIGADA!

Quase me apetece dizer que não há uma Língua Portuguesa. Há línguas em português. É uma língua que tinha de passar, inevitavelmente, por transformações segundo os lugares onde a falam, as culturas e as influências. Mas isso não tira nada da evidência de que trata do corpo da língua portuguesa. É um corpo espalhado pelo mundo. (SARAMAGO, 2012)

RESUMO

Esta pesquisa objetivou investigar o uso da concordância nominal de número em redações produzidas em português por estudantes guineenses e timorenses no âmbito do exame vestibular da Universidade de Integração Internacional da Lusofonia Brasileira (UNILAB), localizada em Redenção-CE. Está embasada por dois principais subsídios teóricos: 1) a Sociolinguística, em sua abordagem variacionista (Labov,[1972] 2008), a qual, a partir da heterogeneidade das línguas naturais, procura analisar as regularidades que podem ser encontradas na variação, investigando variáveis que agem tanto linguística quanto extralinguisticamente, na interseção língua-sociedade; e 2) o trabalho precursor desenvolvido por Scherre (1988), o qual revela que, para a concordância plural, particularmente, a linearidade tem papel fundamental, demonstrando que as ocorrências de concordância no sintagma nominal são interdependentes. O método de abordagem utilizado foi o indutivo e os métodos de procedimento foram o comparativo e o estatístico, com análise quantitativa dos dados e realização de estudo contrastivo entre o português de Guiné-Bissau (PGB) e o português de Timor-Leste (PTL). O estudo contrastivo deste trabalho examinou as variedades não-europeias do português selecionadas, numa perspectiva de análise atomística, utilizando-se algumas das variáveis linguísticas empregadas por Scherre (1988), a saber: a classe gramatical do elemento, sua posição linear no sintagma e a saliência fônica nas três dimensões aplicadas por Scherre (processos morfofonológicos de formação de plural, tonicidade e número de sílabas), associadas às variáveis extralinguísticas que selecionamos: o país de origem (Guiné-Bissau ou Timor-Leste), o sexo (homem ou mulher) e a situação do candidato escrevente no vestibular (aprovado ou reprovado). Observamos aspectos linguísticos e extralinguísticos e sua relevância no uso da concordância de número no sintagma nominal. Resultados globais mostraram que: 1) apesar do contato linguístico do PGB com o crioulo guineense e do PTL com o tétum - línguas consideradas, nesta pesquisa, como maternas dos falantes guineenses e timorenses, por serem, comprovadamente, as línguas mais faladas nos países em questão -, em que não se observam marcas morfológicas explícitas de número em alguns dos elementos no interior do sintagma nominal, no PGB e no PTL a ausência de concordância não é frequente; 2) a variável extralinguística “sexo” não se apresenta como relevante para análise da variação do fenômeno em

questão; e 3) segundo critérios estabelecidos por Labov (2003), acerca da frequência de produtividade com a qual uma regra de variação ocorre, o uso da concordância nominal de número no PGB e no PTL caracterizou uma regra variável.

Palavras-chave: Concordância nominal. Variação. Variedades do português.

ABSTRACT

This research aimed to investigate the phenomenon of variable agreement based on nominal number in its use in essays produced in Portuguese by Guinean and East Timorese students in the college entrance examination of the University of International Integration of Afro-Brazilian Lusophony (UNILAB) located in Redenção-CE. It is based on two main theoretical contributions: 1) Sociolinguistics, in its variational approach (Labov, [1972] 2008), which, based on the heterogeneity of natural languages, seeks to analyze the regularities that can be found in the variation, investigating variables which act both linguistically and extralinguistically, at the intersection of language and society; and 2) the precursor work developed by Scherre (1988), which reveals that, for plural agreement, in particular, linearity plays a fundamental role, demonstrating that the occurrences of agreement in the noun phrase are interdependent from the same dependent variable. The method of approach was the inductive one and the procedure methods were the comparative and the statistical ones, with quantitative analysis of the data and a contrastive study between the Portuguese spoken in Guiné Bissau (PGB) and that spoken in Timor-Leste (PTL). The contrastive study of this work examined the selected non-European varieties of Portuguese from a perspective of atomistic analysis, using some of the linguistic variables used by Scherre (1988), which are: the grammatical class of the element, its linear position in the syntagma and the phonic salience in the three dimensions applied by Scherre (morphological processes of plural formation, tonicity and number of syllables), associated with the extralinguistic variables we selected: the country the students are from (Guiné Bissau or Timor-Leste), sex (man or woman) and the situation of the student in the college entrance exam (approved or disqualified). We observed linguistic and extralinguistic aspects and their relevance in the use of number agreement in the noun phrase. Overall results showed that: 1) in spite of the linguistic contact of the PGB with the Guinean Creole and the PTL with the Tetum - languages considered, in this research, as the mother language of the Guinean and East Timorese speakers, since they are the most widely spoken languages in countries analyzed in which there are no explicit morphological marks of numbers in some of the elements within the noun phrase, PGB and PTL, the absence of agreement is not frequent; 2) the extralinguistic variable "sex" is not relevant for the analysis of the variation of the phenomenon in question; and 3)

according to Labov's (2003) criteria for the productivity frequency with which a variation rule occurs, the use of nominal number agreement in PGB and PTL characterized a variable rule.

Keywords: Nominal agreement. Variation. Portuguese varieties.

LISTA DE QUADROS

Quadro 1 - Padrões de concordância nominal em três variedades africanas do português.....	22
Quadro 2 - Distribuição dos informantes (total = 400 informantes).....	62
Quadro 3 - Quantitativo geral e percentual de constituintes com “+CN” e “-CN” por país e sexo, obtido do programa computacional <i>goldvarb-x</i>	79
Quadro 4 - Variáveis selecionadas para a análise.....	80
Quadro 5 - Pesos relativos das variantes dos grupos de fatores, conforme ordem em que apareceram na rodada do programa computacional <i>goldvarb-x</i>	82
Quadro 6 - Cruzamento dos dados “posição linear” e “situação no vestibular”.....	86
Quadro 7 - Volume de constituintes por país, por número de sílabas e por presença ou ausência de concordância.....	107
Quadro 8 - Pesos relativos da variável saliência fônica na dimensão número de sílabas, por variante e país.....	110
Quadro 9 - Pesos relativos da variável saliência fônica na dimensão número de sílabas, por variante e país.....	111
Quadro 10 - Cruzamento dos dados “país de origem” e “situação no vestibular”, por quantitativo de constituintes analisados e número percentual.....	114
Quadro 11 - Pesos relativos das variantes dos grupos de fatores – país Guiné Bissau.....	115
Quadro 12 - Pesos relativos das variantes dos grupos de fatores - país Timor-Leste.....	116

LISTA DE GRÁFICOS

Gráfico 1 - Percentual de +CN e -CN, tomando por base o país de origem.....	64
Gráfico 2 - Percentual de +CN e -CN, tomando por base a classe gramatical.....	83
Gráfico 3 - Percentual de +CN e -CN, tomando por base a posição linear do constituinte no sintagma.....	85
Gráfico 4 - Percentual de +CN e -CN, tomando por base a saliência fônica - dimensão processos.....	87
Gráfico 5 - Percentual de +CN e -CN, tomando por base a saliência fônica – dimensão tonicidade.....	88
Gráfico 6 - Percentual de +CN e -CN, tomando por base a saliência fônica – dimensão número de sílabas.....	90
Gráfico 7 - Percentual de +CN e -CN, tomando por base o país de origem.....	91
Gráfico 8 - Percentual de +CN e -CN, tomando por base o sexo.....	92
Gráfico 9 - Percentual de +CN e -CN, tomando por base a situação no vestibular.....	94
Gráfico 10 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a classe gramatical.....	96
Gráfico 11 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base a classe gramatical.....	98
Gráfico 12 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a posição linear.....	100
Gráfico 13 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base a posição linear.....	101
Gráfico 14 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a saliência fônica – dimensão processos.....	104
Gráfico 15 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base a saliência fônica – dimensão processos.....	104
Gráfico 16 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a saliência fônica – dimensão tonicidade.....	106
Gráfico 17 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a saliência fônica – dimensão tonicidade.....	106
Gráfico 18 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a saliência fônica – dimensão número de sílabas.....	108

Gráfico 19 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base a saliência fônica – dimensão número de sílabas.....	109
Gráfico 20 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base o sexo.....	110
Gráfico 21 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base o sexo.....	111
Gráfico 22 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a situação no vestibular.....	113
Gráfico 23 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base a situação no vestibular.....	113

SUMÁRIO

1	INTRODUÇÃO.....	18
1.1	NOSSO OBJETO DE ESTUDO.....	19
1.2	OS OBJETIVOS.....	25
1.2.1	Geral.....	25
1.2.2	Específicos.....	25
1.3	AS HIPÓTESES.....	26
1.3.1	Geral.....	26
1.3.2	Específicas.....	26
1.4	BREVES INCURSÕES TEÓRICO-METODOLÓGICAS.....	27
1.4.1	O arcabouço teórico.....	27
1.4.2	O arcabouço metodológico.....	27
1.5	ESTRUTURA DA DISSERTAÇÃO.....	28
2	CONCEITOS BASILARES: SOCIOLINGÜÍSTICA E ESTUDOS DE CONCORDÂNCIA.....	29
2.1	SOBRE A SOCIOLINGÜÍSTICA.....	29
2.2	LÍNGUAS DE HERANÇA E ESTUDOS SOCIOLINGÜÍSTICOS.....	32
2.3	SOBRE A AQUISIÇÃO DA LINGUAGEM EM CONTEXTO MULTILÍNGUE.....	36
2.3.1	O caso do crioulo guineense e do tétum: aquisição de L1.....	37
2.3.1.1	<i>O caso específico da Guiné Bissau.....</i>	38
2.3.1.2	<i>O caso específico do Timor-Leste.....</i>	41
2.3.2	Aquisição versus aprendizagem.....	44
2.4	ESTUDOS DA CONCORDÂNCIA NOMINAL DE NÚMERO EM VARIEDADES DO PORTUGUÊS.....	45
2.4.1	Nas variedades europeia e brasileira.....	45
2.4.2	Nas variedades africanas.....	49
2.4.2.1	<i>Angola e Moçambique.....</i>	50
2.4.2.2	<i>Guiné Bissau.....</i>	51
2.4.2.3	<i>São Tomé e Príncipe.....</i>	56
2.4.3	No Timor-Leste.....	57
3	PROCEDIMENTOS METODOLÓGICOS.....	60
3.1	MÉTODO UTILIZADO.....	60

3.2	LOCAL DA COLETA DOS DADOS.....	61
3.3	SUJEITOS DA PESQUISA, AMOSTRA, E CONSTITUIÇÃO DOS CORPORA.....	62
3.4	SELEÇÃO DAS VARIÁVEIS.....	66
3.4.1	Variáveis dependentes.....	67
3.4.1.1	<i>De número.....</i>	67
3.4.2	variáveis linguísticas para análise da concordância nominal de Número.....	67
3.4.2.1	<i>Posição linear do elemento no sintagma nominal.....</i>	67
3.4.2.2	<i>Classe gramatical do elemento nominal.....</i>	70
3.4.2.3	<i>Saliência fônica – Dimensão processos morfofonológicos de formação do plural.....</i>	72
3.4.2.4	<i>Saliência fônica – Dimensão tonicidade do elemento.....</i>	73
3.4.2.5	<i>Saliência fônica – Dimensão número de sílabas dos itens lexicais singulares.....</i>	73
3.4.3	Variáveis extralinguísticas para análise da concordância nominal de número.....	74
3.4.3.1	<i>País de origem.....</i>	74
3.4.3.2	<i>Sexo.....</i>	75
3.4.3.3	<i>Situação no vestibular.....</i>	77
3.5	TRATAMENTO QUANTITATIVO DOS DADOS.....	78
4	OS DADOS.....	79
4.1	RESULTADOS GLOBAIS.....	79
4.1.1	Variável classe gramatical.....	83
4.1.2	Variável posição linear.....	84
4.1.3	Variável saliência fônica.....	86
4.1.3.1	<i>Variável Saliência Fônica – Dimensão Processos.....</i>	87
4.1.3.2	<i>Variável Saliência Fônica – Dimensão Tonicidade.....</i>	88
4.1.3.3	<i>Variável Saliência Fônica – Dimensão Número de Sílabas.....</i>	89
4.1.4	Variável país de origem.....	90
4.1.5	Variável sexo.....	91
4.1.6	Variável situação no vestibular.....	93
4.2	CONTRASTANDO AS VARIEDADES GUINEENSE E TIMORENSE DO PORTUGUÊS.....	95

4.2.1	Variável classe gramatical.....	96
4.2.2	Variável posição linear.....	100
4.2.3	Variável saliência fônica – Dimensão processos.....	103
4.2.4	Variável saliência fônica – Dimensão tonicidade.....	105
4.2.5	Variável saliência fônica – Dimensão número de sílabas.....	108
4.2.6	Variável sexo.....	110
4.2.7	Variável situação no vestibular.....	112
4.3	QUADROS-RESUMO DOS PESOS RELATIVOS POR GRUPO DE FATORES E POR VARIANTE PARA O USO DO PGB E DO PTL.....	115
5	CONSIDERAÇÕES FINAIS.....	119
	REFERÊNCIAS.....	124

1 INTRODUÇÃO

O português do Brasil (doravante PB) é uma variedade do português já bastante consolidada e difundida, haja vista o processo de gramatização desta variedade ter sido iniciado ainda no século XIX, após a independência de Portugal, o que permitiu seu estudo sistemático a partir de diversas abordagens (FARACO, VIEIRA, 2016). Por outro lado, pelo fato da independência dos países africanos de língua portuguesa ainda ser recente, não se tem, neles, um processo de gramatização consolidado (SEDRINS; SILVA, 2017). Ademais, a existência de gramáticas voltadas para a descrição da língua portuguesa do Brasil, como aquelas postas em perspectiva na obra organizada por Faraco e Vieira (2016), favorece o desenvolvimento de linhas de pesquisas que proporcionam o conhecimento aprofundado da variedade brasileira do português.

Assim, outras variedades não europeias do português também precisam ser estudadas a fim de que se conheça melhor sua realidade linguística. No Brasil, o português é língua oficial a exemplo do que ocorre nos cinco países lusófonos da África (Moçambique, Guiné-Bissau, Angola, Cabo Verde e São Tomé e Príncipe) e no país asiático de Timor-Leste, embora uma distinção se observe entre o Brasil e esses países: naquele, o português é a língua materna da maioria dos falantes, ao passo que nestes, é, em geral, uma segunda língua, convivendo com as línguas maternas dos falantes, o que caracteriza, assim, um intenso multilinguismo.

Vale referir que, para uma maior integração e estabelecimentos de parcerias entre os países que têm o português como língua oficial, foi criada, em 1996, a Comunidade dos Países de Língua Portuguesa (CPLP) que integrou Angola, Brasil, Cabo Verde, Guiné-Bissau, Moçambique, Portugal e São Tomé e Príncipe. Apenas em 2002 o Timor-Leste passou a integrar essa comunidade. Salientamos que, apesar dessa proposta de integração entre os países lusófonos, existe, ainda nos dias atuais, grande lacuna e latente desequilíbrio no prestígio de algumas variedades do português, sobretudo as africanas e a asiática, do Timor-Leste, o que desestimula os investimentos em pesquisas linguísticas relativas a elas, trazendo, como consequência, o já mencionado processo de gramatização tardio e um número limitado de outras pesquisas.

Um exemplo desse problema são os diversos estudos no âmbito da concordância nominal, por exemplo, que têm sido realizados ao longo dos anos no

no PB e no português europeu (doravante PE), o que não se equipara aos estudos relativos às outras variedades. A importância de estudos que tratem de aspectos morfossintáticos de variedades do português mostra-se na possibilidade de compreender os processos que provocam, conforme dito por Brandão (2016), uma “polarização sociolinguística”, a qual, por sua vez, contribui para a formação identitária tanto de países lusófonos da África cuja independência da metrópole portuguesa é ainda recente, quanto do país asiático de Timor-Leste.

1.1 NOSSO OBJETO DE ESTUDO

Esta pesquisa centra a atenção em aspectos morfossintáticos, mais especificamente, no uso da concordância nominal de número em redações produzidas em português por alunos guineenses e timorenses que prestaram exames para ingresso na Universidade de Integração Internacional da Lusofonia Brasileira (UNILAB), localizada em Redenção, no Ceará. Assim, serão analisadas duas variedades não-europeias do português em dois países lusófonos de continentes diferentes: a saber, o português de Guiné Bissau (doravante PGB) e o de Timor-Leste (doravante PTL), considerando seu intenso contexto de multilinguismo. Apesar de não negarmos a existência de outras línguas maternas em Guiné-Bissau e em Timor-Leste, consideraremos, para fins de análise neste estudo, a convivência do PGB com o crioulo guineense, língua falada por mais de 90% das pessoas, considerando todas as etnias do país, segundo dados do Recenseamento Geral da População e Habitação, referente ao ano de 2008 e publicado em 2009. Do mesmo modo, em relação ao PTL, consideraremos o seu contato com o tétum, língua também oficial da nação. Deter-nos-emos, portanto, ao contato linguístico entre as duas variedades do português e essas línguas maternas, tomando por base a teoria da sociolinguística variacionista (LABOV [1972] 2008).

Outro aspecto relevante quando falamos de línguas em contato, é o estatuto que a língua recebe na localidade. No caso da Guiné-Bissau, por um lado, o português é visto como uma segunda língua (L2), apesar de sua função social não ser a de promover integração social, mas de funcionar como língua de poder (PEIXOTO e SOARES, 2014). Por outro lado, no Timor-Leste, embora o português tenha o estatuto de língua oficial, confirmado pelo artigo 13º da Constituição da

República Democrática de Timor-Leste (ALMEIDA, 2008), o que faz dele a língua de instrução/escolarização e língua veicular de acesso aos conhecimentos de outras disciplinas, verifica-se uma situação um pouco mais complexa: nesse país, o estatuto de língua oficial implicaria que o PTL fosse uma segunda língua, visto que ela é necessária para a participação na vida política e econômica da nação, além de ser a língua da escola, no entanto, como o *input* existente no país é insuficiente para que o PTL possa ser aprendido sem o recurso da escola, ele pode ser considerado uma LE (ALMEIDA, 2008, p. 43). O autor afirma que

como se verificou, se se observa o estatuto sociopolítico que o português tem em Timor, pode dizer-se que é PLS¹. Por outro lado, quando é tido em conta o contexto de aprendizagem, mais ou menos formal, parece ser mais proveitoso recorrer ao conceito de PLE², no sentido prático de aplicar estratégias minimizadoras do fraco *input* externo. (ALMEIDA, 2008, p. 43)

Ainda segundo Almeida (2008),

se, relativamente ao contexto mais frequente de ensino formal, a língua portuguesa em Timor-Leste se inclina mais para PLE, visto que esse é inegavelmente o contexto em que os jovens timorenses acedem a esta língua, no que concerne ao facto da LE³ ser aprendida em espaços fisicamente distantes daqueles em que é falada. (ALMEIDA, 2008, p. 44)

Estamos, portanto, diante de duas realidades distintas para o português em Guiné Bissau e em Timor-Leste: nesta, o PTL caracteriza uma LE, enquanto, naquela, o PGB constitui uma L2.

Os estudos sociolinguísticos que vêm sendo desenvolvidos no Brasil, especificamente aqueles que se voltam para análises morfossintáticas, como as referentes à concordância nominal, apontam para o fato de que se trata de uma regra variável nesse país através da convivência da variante padrão com a não-padrão (ex.: os meninos ~ os menino) (SCHERRE, 1988; SCHERRE; NARO, 1998), ao contrário do que se observa em Portugal, que se apresenta como uma regra categórica em sua aplicação (ex: as casas) (BRANDÃO; VIERA, 2012a). Portanto, cabe, nesta pesquisa, investigar se os dados do PGB e do PTL se aproximam mais da variedade europeia do português ou se já demonstram características próprias

¹ PLS: Português como Língua Segunda.

² PLE: Português como Língua Estrangeira.

³ LE: Língua Estrangeira.

oriundas do contexto de multilinguismo existente nesses países, tendo em mente que a situação de contato do português com uma das línguas maternas em Guiné Bissau e no Timor-Leste pode exercer grande influência no mecanismo dessa concordância.

Ademais, estudos revelam que, em variedades não-europeias do português, o grau de escolaridade exerce influência no uso linguístico no sentido de que as estruturas produzidas por falantes mais escolarizados estão mais próximas da norma europeia do português (BACELAR DO NASCIMENTO et al., 2008; MOTA, MIGUEL, MENDES, 2012), e a organização escolar tanto da Guiné Bissau quanto do Timor-leste favorece tal afirmação, uma vez que o português é a língua de ensino em ambos os países, adotando como modelo a norma europeia do português considerada referência do “bom português”.

Conforme nota publicada pela UNESCO (JOUDE, 2016), em Guiné Bissau, 45% das crianças, ou seja, aproximadamente metade delas, em idade de escolarização obrigatória encontram-se fora da escola, realidade atribuída à pouca disponibilidade de escolas que ofereçam todos os anos do ensino regular, como Ensino Fundamental e Médio se compararmos ao sistema educacional brasileiro. Considerando o fato de o crioulo guineense ser a língua que predomina no uso cotidiano desse país e de o PGB ser mais usado em contextos escolares, essa situação é particularmente relevante, pois nos permite entender como o pouco contato dos habitantes com o PGB, devido à fragilidade na educação formal, pode influenciar no entendimento e utilização de marcas formais de número na concordância nominal.

Pesquisas acerca de outras variedades do português que não sejam a europeia e a brasileira, fornecem significativa contribuição para o estudo de suas particularidades nos domínios fonético-fonológico, morfossintático, léxico-semântico, entre outros. Os exemplos de (1) a (7), abaixo, constituem dados de fala produzidos por falantes escolarizados de Moçambique (M) e evidenciam a relação entre escolaridade e uso da norma-padrão do português:

(1) “não foi possível continuar os *estudos* [...]” (M)

(2) “tivemos debates com *certos professores* e *certas visitas* [...]” (M)

(3) “os meus amigos estão mais *adiantados* em termos dos estudos [...]” (M)

(4) “os meus dois *dias* no caminho [...]” (M)

(5) “não tive assim *muitos problemas* [...]” (M)

(6) “diferença quanto aos *nossos transportes* de cá [...]” (M)

(7) “chapas tipo *táxis* – - e *transportes públicos* [...]” (M)

(BACELAR DO NASCIMENTO et al., 2008, p. 375)

Enquanto o grau de escolaridade, quando elevado, aproxima a estrutura das variedades africanas do português da norma europeia, devido à influência do PE no processo de escolarização, e isto se refletir nas construções de falantes de variedades africanas do português, o baixo grau de escolaridade faz com que os falantes produzam sintagmas nominais com ausência de concordância de número. A Tabela 1, abaixo, extraída de Brandão (2016, p.89), nos apresenta dados de fala de pessoas com poucos anos de escolarização. Nela, embora haja sentenças⁴ em que a regra de concordância foi aplicada, o baixo nível de escolaridade apresenta-se saliente para a manutenção do padrão da não concordância:

Quadro 1 - Padrões de concordância nominal em três variedades africanas do português⁵

PA	(a) quais são [os principais problemas] (b) [as coisa] estão muito cara (c) e buscar [as tuas fruta] (d) [os tais português], fomos [...]
PM	(a) graças a[os meus irmão mais velhos] (b) tinha que levar [as criança]

⁴ Na sentença (a) tanto do PA quanto do PST, observamos o uso da concordância de acordo com a norma culta. Apenas dois casos da aplicação da regra entre as treze sentenças expressas na tabela de Brandão (2016, p. 89). Uma minoria, cuja razão podemos atribuir à baixa escolarização dos usuários.

⁵ PA: Português de Angola; PM: Português de Moçambique; PST: Português de São Tomé e Príncipe.

	(c) [alguns nossos amigo] vê uma [...] (d) e d[os acontecimentos alegre] (e) boda de prata de [vinte e cinco ano]
PST	(a) n[as antigas empresas coloniais] (b) morre [aquelas lula mesmo grande] (c) não era n[os nossos tempo] (d) pego n[as minha ferramenta]

Fonte: *Corpus Variedades Africanas do Português (VAPOR)*, do Centro de Linguística da Universidade de Lisboa. (BRANDÃO, 2016, p.89)

A partir da leitura de trabalhos realizados sobre a concordância de número no sintagma nominal no PB (SCHERRE, 1988; SCHERRE; NARO, 1998, 2006), que concluem ser um fenômeno variável nessa variedade e ser influenciado por variáveis linguísticas e extralinguísticas, ampliamos, nesta pesquisa, a discussão a partir de um estudo comparativo entre duas variedades não europeias do português, a saber: o PGB e o PTL, cujas pesquisas sociolinguísticas no âmbito da concordância nominal ainda são escassas até onde temos verificado. Além disso, a perspectiva de análise contrastiva aqui proposta nesta dissertação é pioneira no âmbito desses estudos.

Quando falamos em “fenômeno variável” referimo-nos à caracterização dos padrões de variação proposta por Labov (2003, p. 243), o qual nos apresenta critérios para que classifiquemos a variação sociolinguística. Tais critérios podem ser assim descritos: categóricos, para regras com 100% de frequência na produtividade; semicategóricos, para regras que variem de 95 a 99% na frequência de produtividade; e variáveis, para regras de variação com frequência de 5 a 95% de produtividade.

A respeito da frequência com a qual padrões de variação no âmbito da concordância nominal são observados, Brandão (2016, p. 92, 95-96), testa a hipótese de Labov (2003, p. 243) e nos apresenta os seguintes dados quantitativos: no PE, a regra mostra-se categórica, com uma oscilação entre 99,78% e 99,96% de sua aplicação; no PB, trata-se de uma regra variável, com índices que variam entre 44% (*Corpus APERJ*) e 91,6% (*Corpus Concordância*) de sua aplicação; e, no português de São Tomé e Príncipe (doravante PST), a regra também se apresenta como variável, com uma marca e 93,4%.

A hipótese de que, no PGB e no PTL, a concordância nominal se apresenta como uma regra variável contradiz o que se espera na modalidade que está sendo analisada: a língua escrita, uma vez que, em tal modalidade, é comum haver um maior grau de monitoramento no uso da língua. Isso evidencia outro aspecto relevante desta pesquisa: ao contrário de muitos estudos sociolinguísticos que se voltam à língua falada, este estudo centra a atenção na língua escrita, considerando que o registro formal da elaboração de uma redação feita durante o processo de ingresso em uma instituição de Ensino Superior exige um grau de monitoramento maior dos alunos, o que pode nos revelar usos bastante próximos da norma europeia do português. Em caso de alguns usos fugirem a essa norma, aventamos a hipótese da influência da língua materna dos falantes sobre o PGB e o PST.

Apesar de estudos sociolinguísticos enfocarem bastante a língua falada, encontramos pesquisas, embora ainda escassas, sobre a concordância nominal na língua escrita. Silva (2017), realizando um estudo sobre a concordância nominal de número na língua escrita de falantes da cidade de Belo Jardim, localizada no Estado de Pernambuco, fez um levantamento dos estudos sociolinguísticos realizados em outros estados brasileiros e verificou apenas dois: o de Christino e Silva, realizado em 2012, no estado do Rio Grande do Sul, e o de Mariano, desenvolvido em 2013, no Rio de Janeiro. Sobre o PB, verificamos ainda a escassez de pesquisas com a modalidade escrita. Já, no que se refere ao PGB e ao PTL, há escassez de trabalhos não só na modalidade oral, mas também na modalidade escrita. Nesta dissertação, decidimos optar pela modalidade escrita por razões associadas ao grau de monitoramento do uso da língua, entendendo que a situação de escrita de uma redação em um exame vestibular pode influenciar na aplicação de regras linguísticas como as do âmbito da concordância nominal, fenômeno analisado nesta investigação. Outra razão que motivou o estudo da modalidade escrita da língua foi o acesso que tivemos ao banco de dados de textos escritos produzidos por estudantes guineenses e timorenses na ocasião do exame vestibular da Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB).

De maneira geral, este trabalho contribui com os estudos morfossintáticos da concordância no sintagma nominal, uma vez que se trata de um estudo contrastivo entre variedades do português. Em adição, enfatiza-se que o estudo sociolinguístico variacionista no contexto dos países escolhidos fornece suporte empírico para que

se combatam ideologias preconceituosas e excludentes justificadas pelas diferenças linguísticas.

1.2 OS OBJETIVOS

1.2.1 Geral

Investigar o uso da concordância nominal de número em redações produzidas em português por estudantes guineenses e timorenses no âmbito do exame vestibular da UNILAB-CE.

1.2.2 Específicos

- (a) Descrever a possível interferência da morfossintaxe do tétum e do guineense no PTL e no PGB, respectivamente, em particular, no âmbito da concordância nominal de número, tendo em mente a aquisição do português como L2 e sua coexistência com as línguas maternas (L1) desses países;
- (b) Contrastar a morfologia flexional de número no crioulo guineense e no tétum com a do PGB e do PTL encontrada nos dados em análise, tendo em mente possíveis interferências da L1 na L2;
- (c) Verificar fatores de natureza linguística e extralinguística que podem favorecer o uso da concordância nominal de número no PGB e no PTL;
- (d) Analisar os padrões de concordância nominal de número verificados no PGB e no PLT, levando em conta os tipos de regras linguísticas propostas por Labov (2003) à luz da Sociolinguística Variacionista.

1.3 AS HIPÓTESES

1.3.1 Geral

Embora sejam o crioulo guineense e o tétum as línguas maternas dos falantes guineenses e timorenses, respectivamente, em que não se observa, na maioria dos usos, marcas explícitas de número em alguns dos elementos no interior do sintagma nominal (CASTRO, 2013; ALBUQUERQUE, 2012), seria possível que, no PGB e no PTL, a ausência de concordância não fosse tão frequente nos dados escritos analisados, por duas razões: 1) o fato de tratar-se de um contexto monitorado: a situação de um exame; e 2) a influência do PE.

1.3.2 Específicas

- (a) O contato linguístico em Guiné-Bissau e Timor Leste, a aquisição do português como L2 e sua coexistência com as línguas maternas (L1) desses países, respectivamente, o crioulo guineense e o tétum, fariam com que a aplicação da regra de concordância seja um fenômeno variável;
- (b) A morfologia flexional de número no crioulo guineense e no tétum com a do PGB e do PTL permitiria a identificação de possíveis interferências da L1 na L2;
- (c) Fatores de natureza linguística e extralinguística poderiam favorecer o uso da concordância nominal de número no PGB e no PTL;
- (d) Evidências encontradas na análise sociolinguística poderiam comprovar que o tipo de regra que se aplica aos padrões de concordância de número no PGB e no PTL possuiriam característica de uma regra variável.

1.4 BREVES INCURSÕES TEÓRICO-METODOLÓGICAS

1.4.1 O arcabouço teórico

Esta dissertação está embasada na Sociolinguística, em sua abordagem variacionista (LABOV, [1972] 2008), a qual oferece os instrumentos necessários para o estudo da língua em seu contexto social, considerando sua dinamicidade através da variação que lhe é inerente e que, por sua vez, pode ser observada e descrita. Dentro dessa abordagem, retomaremos, dentre tantos trabalhos, o trabalho precursor desenvolvido por Scherre (1988) para análise do fenômeno variável da concordância de número no sintagma nominal do PB.

Diante do exposto, pretendemos desenvolver uma investigação pormenorizada sobre a concordância nominal de número no PGB e no PTL evidenciando que variáveis linguísticas e extralinguísticas podem favorecer determinados padrões de concordância sejam aqueles associados à variante padrão ou não.

1.4.2 O arcabouço metodológico

Esta pesquisa foi empreendida adotando-se o método de abordagem indutivo e os métodos de procedimento: comparativo (entre Guiné Bissau e Timor Leste) e estatístico. Dessa forma, foi realizada a análise quantitativa dos dados e o estudo contrastivo entre o PGB e o PTL, variedades não-europeias do português selecionadas para compor o escopo deste trabalho, as quais foram investigadas no que se refere à concordância de número no sintagma nominal.

Os *corpora* são constituídos de sintagmas nominais extraídos das redações coletadas em que há a variante padrão e não-padrão relacionada à concordância nominal de número. Concluída essa etapa, as variáveis linguísticas e extralinguísticas foram selecionadas, tomando por base variáveis já investigadas em pesquisas sobre essa concordância em outras variedades do português para fins de futura comparação de resultados. Finalizada essa seleção, os dados foram codificados consoante aos fatores que compõem essas variáveis para posterior

rodada dos dados no programa *GoldVarb-X*, seguindo a perspectiva da Sociolinguística Quantitativa.

1.5 ESTRUTURA DA DISSERTAÇÃO

Após a discussão introdutória neste capítulo, com as pontuações iniciais sobre pesquisas já existentes acerca do fenômeno da variação da concordância nominal, as questões que norteiam esta pesquisa, as hipóteses e as breves incursões a respeito das teorias aqui adotadas e da metodologia desta pesquisa, expomos a estrutura do presente trabalho e os aspectos que serão abordados nos capítulos que seguem.

Inicialmente, pontuamos que esta dissertação está dividida da seguinte forma:

a) no capítulo 1, estão expostos a delimitação do objeto de estudo e alguns apontamentos sobre a linguística de contato em contextos de bilinguismo/multilinguismo, a apresentação de evidências do fenômeno em análise, os objetivos, as hipóteses e a estrutura desta investigação; no capítulo 2, discorreremos sobre o quadro teórico adotado, a Sociolinguística, bem como sobre as concepções de bilinguismo e aquisição de língua assumidas; o capítulo 3 trata dos procedimentos metodológicos utilizados nesta investigação. Já o capítulo 4 volta-se à análise linguístico-quantitativa dos dados do PGB e do PTL. Por fim, apresentamos as considerações finais, em que são sintetizados os resultados obtidos ao longo desta dissertação. Deixamos, também, algumas questões sugeridas para eventuais pesquisas futuras que possam ser desenvolvidas partindo de aspectos deixados em aberto neste trabalho.

2 CONCEITOS BASILARES: SOCIOLINGÜÍSTICA E ESTUDOS DE CONCORDÂNCIA

2.1 SOBRE A SOCIOLINGÜÍSTICA

O estudo da língua em seu contexto real de uso, tomando por base a variação que lhe é inerente, constitui uma das áreas da linguística denominada Sociolinguística:

Essa área também atende por outros nomes: (i) Sociolinguística Laboviana, porque seu principal expoente é o linguista norte-americano William Labov; (ii) Sociolinguística Quantitativa, porque, a princípio, os pesquisadores dessa área costumam lidar com uma grande quantidade de dados de usos da língua, o que requer normalmente uma análise estatística; e (iii) Teoria da Variação e Mudança Linguística, por conta de suas principais preocupações: a variação e a mudança na língua. (COELHO et al., 2015, p. 14).

Investigações linguísticas pelo viés da sociolinguística variacionista justificam-se pelo fato de as línguas serem essencialmente dinâmicas e heterogêneas, o que favorece a variação, a qual, por sua vez, pode ser observada e descrita, considerando-se diversos fatores. A heterogeneidade das línguas naturais, nesse cenário, indica que existem diversas maneiras (variantes) de dizermos um mesmo enunciado em um mesmo contexto e com o mesmo valor de verdade (TARALLO, 1985). Essas variantes são escolhidas pelos usuários a partir de motivações diversas tanto na língua falada quanto na língua escrita. Nesse sentido, o sociolinguista procura analisar as regularidades que podem ser encontradas na variação, a partir da investigação de variáveis linguísticas e extralinguísticas que podem estar favorecendo o uso de uma determinada variante.

Labov ([1972] 2008, p. 140) salienta que “os procedimentos da linguística descritiva se baseiam na concepção da língua como um conjunto estruturado de normas sociais”, e que “muitos elementos da estrutura linguística estão envolvidos em variação sistemática que reflete tanto a mudança temporal quanto os processos sociais extralinguísticos”. Assim, a Sociolinguística interessa-se por questões relacionadas à variação e à mudança linguísticas, bem como o que pode ocasionar

estar na gênese desses processos, como por exemplo, o contato entre línguas e o multilinguismo.

Ademais, na perspectiva da Sociolinguística Variacionista, investiga-se o quanto estáveis ou mutáveis são as variações dentro de uma mesma língua ou variedade dela, de modo que seja possível perceber quais variáveis agem positiva ou negativamente sobre os usos, tornando possível a observação de possíveis regularidades. Por variedade, entenda-se “a fala característica de determinado grupo” (COELHO et al., 2015, p. 14), o que coloca, por exemplo, o português brasileiro (PB), o europeu (PE), o de Guiné Bissau (PGB) e o de Timor Leste (PTL) como variedades distintas. Variação, por sua vez, caracteriza a alternância entre duas formas diferentes, mas usadas com o mesmo propósito (COELHO et al, 2015, p. 16), como é o caso das formas “nós” e “a gente”, ambas coexistentes no PB e usadas para se referir à primeira pessoa do plural.

Em estudos de caráter sociolinguístico-variacionista, assume-se que variação linguística não é algo que ocorre aleatoriamente. Ao contrário, defende-se que ela é regida por aspectos internos e externos à língua (linguísticos e extralinguísticos), os quais são chamados de condicionadores, os quais são compartilhados pelos falantes de uma mesma comunidade de fala, isto é, um grupo de pessoas que demonstram certas regularidades/padrões no que pode ser observado acerca da variação linguística existente no local. Segundo Labov ([1972] 2008, p.150),

a comunidade de fala não é definida por nenhuma concordância marcada pelo uso de elementos linguísticos, mas sim pela participação num conjunto de normas compartilhadas; estas normas podem ser observadas em tipos de comportamento avaliativo explícito e pela uniformidade de padrões abstratos de variação que são invariantes no tocante a níveis particulares de uso. (LABOV, [1972] 2008, p.150)

De acordo com Mollica (2013) e Silva (2013), variáveis internas de natureza fonomorfológica, semântica, discursiva e lexical, bem como os diversos subsistemas de uma língua; e variáveis externas inerentes ao indivíduo (como etnia e sexo), propriamente sociais (como escolarização, nível de renda, profissão e classe social) e contextuais (grau de formalidade e tensão discursiva) têm sido investigadas em estudos sociolinguísticos.

Como condicionadores linguísticos, por exemplo, podem ser selecionados para o estudo do fenômeno variável da concordância nominal de número os seguintes fatores: saliência fônica, classe gramatical do elemento nominal, posição linear do elemento no sintagma nominal, marcas precedentes e número de sílabas dos itens lexicais singulares (SCHERRE, 1988); e como condicionadores extralinguísticos podemos ter, por exemplo, o sexo/gênero dos usuários da língua e o grau de escolaridade. Em termos mais específicos, os condicionadores também podem ser chamados de variáveis, havendo a) uma variável dependente que é composta pelas variantes que compõem o fenômeno linguístico variável (por exemplo, a variável dependente relacionada à “expressão pronominal de primeira pessoa do plural” é composta das variantes “nós” e “a gente”) e algumas independentes que constituem os os grupos de fatores linguísticos e extralinguísticos que podem favorecer ou não o uso de uma das variantes.

Metodologicamente, uma investigação pelo viés da sociolinguística variacionista resume-se da seguinte maneira, segundo Coelho et al. (2015, p. 21-22):

identifica-se uma variável no uso corrente da língua de uma comunidade; identificam-se, a seguir, as variantes dessa variável; a partir das hipóteses que elaboramos quanto aos condicionadores que possam estar em jogo no favorecimento ou desfavorecimento das variantes, prossegue-se à coleta de dados reais de fala e, posteriormente, à sua análise, para confirmação ou refutação das hipóteses iniciais. (COELHO *et al.*, 2015, p. 21-22)

Os autores assumem, ainda, que a “natureza do sistema é probabilística, o que pressupõe o emprego de técnicas quantitativas para a observação das regularidades que o regem”. De acordo com Labov ([1972] 2008, p.150),

os indicadores linguísticos oferecem um grande volume de dados quantitativos que refletem a influência de diversas variáveis independentes. Uma vez determinada a significação social de dada variante linguística, [...], esta variável pode então servir de indicador para medir outras formas de comportamento social: aspirações à ascensão social, mobilidade e insegurança social, mudanças na estratificação e segregação social. (LABOV, [1972] 2008, p.150)

Nesta pesquisa, o estudo do português em duas de suas variedades não europeias (PGB e PTL), numa abordagem sociolinguística, possibilita, além de uma melhor compreensão do fenômeno linguístico em análise e das características dessas variedades, o suporte empírico para que se combatam ideologias preconceituosas e excludentes justificadas pelas diferenças linguísticas, sobretudo, por se tratarem de dois países com realidades linguísticas ainda pouco investigadas.

A concordância nominal, fenômeno linguístico deste estudo, apresenta-se como um aspecto da morfossintaxe da língua que pode revelar uma dinâmica sociolinguística bastante diversificada quando levadas em conta diferentes variedades do português, em nosso caso, o PGB e o PTL a partir de suas diferentes histórias de contato linguístico.

2.2 LÍNGUAS DE HERANÇA E ESTUDOS SOCIOLINGUÍSTICOS

Segundo Lima-Hernandes (2016), o conceito para língua de herança, atualmente, engloba os estudos relativos a línguas em contato, língua materna, bilinguismo circunstancial, manutenção/preservação de línguas, L2, língua doméstica, desaparecimento de língua entre outros, o que nos permite analisar contextos peculiares de uso de determinada língua em locais onde, embora haja uma língua dada como nacional/oficial, os falantes utilizam, no ambiente doméstico, outra(s) língua(s). Nessa perspectiva, a noção de língua como representação de uma identidade social/cultural atrela-se às discussões acerca de políticas linguísticas e à noção de superioridade e/ou inferioridade entre línguas.

Outro aspecto relevante acerca das línguas de herança é que, comumente, em regiões onde há disparidade entre língua oficial, ensino e língua com a qual os falantes associam sua identidade, como é o caso de alguns dos países africanos de língua portuguesa e do Timor Leste, a língua majoritária tende a ser aquela falada pela maioria dos usuários ou por aqueles que possuem o estatuto social de poder, que acaba sendo evidenciado quando se trata de contextos de colonização. Essa característica coloca as línguas de herança, as quais tendem a ser transmitidas entre as gerações através da oralidade, em um lugar “menor”, de pouco uso e estigmatizadas socialmente. Associado a isso está, por exemplo, o fato de os estudos sobre as línguas crioulas faladas em países lusófonos serem ainda

escassos até onde temos verificado e, por serem tais línguas, predominantemente, orais, vão perdendo espaço para o português, cujos estudos são abundantes, sobretudo, em suas variedades europeia e brasileira.

Nesse cenário tão peculiar que envolve as línguas de herança insere-se o contexto sociolinguístico da Guiné Bissau e do Timor Leste: nesses países, a língua oficial natural do país não é a mais falada e o fenômeno do *codeswitching*, bem como as interferências fonológicas, morfossintáticas e lexicais são frequentes.

Os contatos linguísticos entre línguas de herança e língua oficial promovem, geralmente, a existência de indivíduos bilíngues simultâneos, os quais possuem duas línguas maternas⁶. No entanto, no contexto dos países investigados neste estudo, essa característica não se aplica. Em Guiné Bissau e em Timor-Leste, o português é aprendido, basicamente, na escola, o que faz com que poucos sejam os indivíduos falantes. As variedades do português faladas nos referidos países constituem-se como segunda língua dos indivíduos, sendo a capacidade de interagir de maneira fluente, muitas vezes, associada aos anos de escolarização. Isso evidencia que, mesmo em situações de contato linguístico com línguas de herança, muito se perde pela baixa quantidade no número de usuários. Na Guiné Bissau, por exemplo, esse contato acontece entre o crioulo guineense e o português, além de outras línguas étnicas, de herança. O mesmo acontece no Timor-Leste, onde a língua predominantemente falada e com a qual a população se identifica é o tétum, enquanto o português perdura como idioma oficial do país. Em ambos os países, não vemos, portanto, situação equilibrada de bilinguismo.

Lima-Hernandes (2016, p. 101) nos diz que “nesses contatos, riquezas são perdidas por imposição ou por livre busca de enquadramento social”, uma vez que, por decisão dos pais, no período de aquisição de língua por uma criança, é dada a preferência para a língua oficial, ainda que ela não seja a majoritária do ponto de vista de ser falada pela maioria da comunidade, mas, sim, por ser utilizada por aqueles que gozam de maior prestígio social. Segundo a autora,

como o estatuto oficial de línguas é determinado por critérios essencialmente políticos, nesses espaços as soluções encontradas para eleição da língua oficial foram norteadas pelas circunstâncias localizadas e,

⁶ Assumimos com Lima-Hernandes (2016, apud MOLLICA e FERRAREZI JÚNIOR, 2016), que a língua materna nem sempre se trata da língua da mãe, mas, sim, daquela adquirida em ambientes domésticos e quando o indivíduo ainda estava na primeira infância.

por isso, são bem diversas. Essas políticas podem, a longo prazo, afetar drasticamente as línguas maternas, e podem tornar o que era materna numa língua de herança.

(LIMA-HERNANDES, 2016, p. 101)

Lima-Hernandes (Ibid.) nos diz, ainda, que “do ponto de vista de uma língua de herança, deve-se considerar o valor cultural agregado ao indivíduo bilíngue, o que demandará políticas linguísticas de preservação como um direito social e político”.

Sobre a relação entre língua materna e identidade do indivíduo, Revuz (1998), atrela a aquisição de uma segunda língua à perda de carga afetiva, uma vez que nomeávamos as coisas e os seres em nossa língua mãe e passamos a operar a nomenclatura numa língua estrangeira. Segundo a autora, há um deslocamento de marcas anteriores e as novas unidades de significação são desprovidas de cargas afetivas. Em outras palavras, a língua estrangeira não seria capaz de recortar o real do mesmo modo que a língua materna o faz, devendo o falante, portanto, adotá-la para permitir que tais interferências de significação tornem-se amalgamadas a cargas de afetividade.

Esse processo de não reconhecer plenamente o significado de uma palavra na língua lexificadora, no sentido de conhecê-la/sabê-la, mas não senti-la, culminaria, conforme Revuz (Ibid.), na ilusão de que traduzir termos e procurar correspondências entre coisas e palavras fosse suficiente para abranger todas as significações possíveis em uma língua. Para a autora, a aquisição de uma segunda língua levaria, inicialmente, a um estranhamento provocado pela utilização de palavras apenas parcialmente contaminadas pelos valores presentes nas palavras da língua materna. Tal estranhamento poderia representar, simultaneamente, uma perda (até mesmo de identidade), embora promovesse uma renovação ao operar nomeações.

O contato linguístico entre línguas de herança e outras línguas oficiais majoritárias relaciona-se ao que é defendido por Revuz (Ibid.), uma vez que, durante todo esse processo, a língua de superestrato acaba por ocupar um espaço maior na utilização, embora essa realidade não se apresente no PGB e no PTL. Nesses países, o português ainda é pouco falado pela população e, em geral, só é aprendido na escola. pelos falantes de determinada comunidade. Isso não significa,

necessariamente, que a língua majoritária tomará o lugar da língua de substrato e passará a ser usada indiscriminadamente, mas, sim, que ela contribuirá com o surgimento de uma terceira língua, nova, formada por um processo de pidginização/crioulização, o que, por sua vez, acaba por transformar uma língua de substrato em uma língua de herança. Isso porque, na ocasião da gênese de uma língua crioula, partindo de um *pidgin*, a criança em período de aquisição passa a ter outras percepções, outros conceitos, e a produzir enunciados sobre referentes já conhecidos na língua materna, mas que nem sempre são correspondidos plenamente na língua de superestrato, o que acarreta em uma reestruturação léxico-gramatical e conseqüente desenvolvimento do *pidgin*, que acaba evoluindo para crioulo.

A existência das línguas de herança possui raízes também nos processos europeus de colonização e na realidade do período de escravidão. Essas raízes justificam a perda de características estruturais/gramaticais e sociais das línguas maternas das pessoas escravizadas, fato que acaba influenciando na escolha dos pais em desejarem que seus filhos adquiram a língua dos grupos dominantes, pois eles costumam vir de experiências traumáticas de quem precisou deixar seu lugar/país de origem para viver em outro país e, de imediato, adaptar-se àquela cultura e àquela língua. Nesse contexto, quando os pais de crianças da comunidade em que se fala a língua de substrato decidem por oferecer aos seus filhos a oportunidade de adquirirem a língua do grupo dominante, eles esperam que estes desenvolvam o sentimento de pertencimento àquele grupo dominador, àquela cultura de distanciamento de um povo escravizado e, portanto, marginalizado. Nesse contexto, línguas maternas de falantes adultos de um grupo dominado tornam-se línguas de substrato e de herança quando as crianças da comunidade já não as adquirem como língua primeira, como é o caso das línguas africanas que estão na formação do crioulo guineense.

Em suma, podemos dizer que adquirir uma segunda língua em situação de intenso e prolongado contato linguístico com a língua materna favorece o surgimento de pidgins e crioulos, o que, conseqüentemente, acaba por colocar as línguas de substrato na posição de língua de herança. Além disso, a variedade de línguas de herança propicia uma situação de pidginização/crioulização repleta de variações linguísticas sistemáticas, observáveis e descritíveis, aspectos sobre os quais a sociolinguística variacionista se debruça.

Considerando os aspectos aqui abordados sobre a constituição de uma língua crioula, voltamos nossa atenção à constituição do PGB e do PTL, uma vez que essas variedades do português foram formadas por processos de pidginização e criouliização derivados do contato linguístico com diversas línguas étnicas existentes na Guiné Bissau e no Timor-Leste, muitas delas de herança.

2.3 SOBRE A AQUISIÇÃO DA LINGUAGEM EM CONTEXTO MULTILÍNGUE

O multilinguismo é um fato bastante relevante acerca da comunicação, interação entre comunidades de fala ao redor do mundo. A existência de línguas francas, as quais facilitavam o entendimento entre esses grupos representa, há bastante tempo, o prestígio de determinado grupo em detrimento de outros na sociedade. Nesse cenário, temos, por exemplo, o latim, o qual foi, durante muito tempo, uma língua franca, assim como o inglês se apresenta nesse lugar atualmente. As línguas pidgins, que se formam nesses contextos multilíngues, também caracterizam “espécies de línguas francas” que surgem para viabilizar a comunicação entre diferentes grupos étnicos que convivem uma mesma localidade (EDWARDS, 1994).

Tanto na Guiné Bissau quanto no Timor-Leste, países investigados neste estudo, há a situação de multilinguismo devido à existência de diversos grupos étnicos vivendo em uma mesma localidade. Além disso, o processo de colonização portuguesa, pelo qual ambos os países passaram, colocou a língua portuguesa em contato com as diversas línguas naturais desses grupos, caracterizando o PGB e o PTL como variedades do português que se diferenciam do PE em vários aspectos.

No contexto de colonização e de inferiorização de um país por outro, a circunstância de ser multilíngue pode ser considerada como algo ruim, no entanto, essa característica dos países aqui estudados, bem como de outros países lusófonos que também se apresentam em situação de multilinguismo, demonstra a necessidade de muitos povos no mundo atualmente, tornando tal situação bastante comum tanto na realidade dos países investigados como de outros, o que deve colocar essa característica dentro do que se entende como normalidade. Quando, na situação de multilinguismo de uma comunidade, há a consequência de que uma

determinada língua se sobreponha a outra e que abranja, dessa forma, maiores possibilidades de comunicação, elevando-a ao estatuto de oficial, como é o caso das variedades do português no contexto da Guiné Bissau e do Timor-Leste, isso pode acarretar a diminuição das línguas naturais dessas localidades e consequente perdas culturais.

Segundo Edwards (1994), há, no mundo, cerca de 5000 línguas faladas em, aproximadamente, 200 países. Esse fato, em sua gênese, já favorece o multilinguismo, visto que esses números constituem, em média, 25 línguas distintas sendo faladas em cada país. Nos países africanos, por exemplo, existem comunidades tão complexas linguisticamente que a situação de multilinguismo é uma realidade com a qual se convive comumente. Na Nigéria, por exemplo, o inglês figura como língua oficial, mas cerca de 80 milhões de pessoas falam algo em torno de 400 línguas, devido aos padrões de interação social. Nesse cenário, Edwards (1994), defende que a diversidade linguística é um componente da vida humana.

2.3.1 O caso do crioulo guineense e do tétum: aquisição de L1

Grande parte dos *pidgins* e dos crioulos formou-se durante o período da expansão marítima que favoreceu os processos europeus de colonização. O estudo dessas línguas, resultantes do contato linguístico, é subsidiado por um ramo da linguística denominado Crioulística, a qual estuda do que são formadas e quando/como se originaram.

Línguas originadas por processos de pidginização e crioulização costumam herdar o léxico da língua de superestrato, embora - inicialmente, em sua fase de pré-pidgin - possuam uma estrutura gramatical diferente e bastante simplificada em relação à língua lexificadora. Lucchesi (2016) observa que

as línguas pidgins e crioulas resultam do processo de estruturação gramatical do código de comunicação emergencial que os falantes utilizam para a comunicação interétnica no momento inicial do contato (o jargão ou pré-pidgin), que consiste de um reduzido elenco de itens lexicais utilizados para formar frases praticamente desprovidas de elementos gramaticais. (LUCCHESI, 2016, p. 73)

A reestruturação gramatical de um pré-pidgin, como afirma Lucchesi, amplia suas funcionalidades e dá origem a *pidgins* estáveis, também chamados de expandidos. Isso porque, em fases iniciais do contato linguístico, os usuários utilizam a “nova” língua em situações de submissão de grupos falantes da língua de substrato, como receber ordens e outras trocas de informações básicas. No cenário da pidginização, falantes adultos utilizam a “nova língua” para se comunicarem e o pidgin não é sua língua materna. Quando, no entanto, as crianças do grupo dominado começam a adquirir a língua *pidgin* como primeira língua, consolida-se o processo de crioulização.

Lucchesi (Ibid, p. 74), ao assumir a relação entre crioulização e o colonialismo europeu, constituída por relações de submissão, levanta três cenários possíveis nos quais esse processo se dá, havendo: a) o crioulo de plantação, quando a crioulização ocorria em torno das grandes empresas agroexportadoras; b) crioulo de quilombo, quando se originava em agrupamentos de escravos foragidos; e c) crioulo de forte, quando a crioulização era favorecida por entrepostos comerciais e/ou fortificações estabelecidas pelos europeus ao longo da costa africana e em outros lugares.

Dado o processo de aquisição de linguagem numa situação de contato linguístico que resultou em crioulização, assumimos com Lucchesi (Ibid.) que “os crioulos não são línguas distintas das demais línguas humanas, pois também são línguas maternas de seus falantes, gozam de plenitude funcional e estruturam-se de acordo com os princípios da faculdade da linguagem proposto por Chomsky”. E ainda “a estruturação gramatical do crioulo seria o produto da ação direta do dispositivo inato de aquisição da língua materna, comum a todos os seres humanos, e parte do patrimônio genético da espécie” (Ibid.).

2.3.1.1 O caso específico da Guiné Bissau

Segundo Namone e Timbane (2017), vários fatores podem nos explicar porque o uso do português não é tão expressivo em relação ao crioulo guineense. O primeiro é que a região era lugar de trânsito de escravos e as pessoas não chegavam a formar famílias sólidas. Assim, os poucos que ficavam aprendiam o

crioulo como língua franca nas suas inter-relações, para aquilo que eram suas necessidades. Além disso, o fato de a colonização portuguesa ter a Guiné-Bissau como colônia de exploração, e não de ocupação, não permitiu a expansão da língua portuguesa para todas as regiões do território, limitando-se apenas ao litoral e nas capitais. Portanto, os colonizadores portugueses usavam o português como instrumento de dominação, e aos dominadores eram dadas instruções básicas, como ler, escrever e falar corretamente o português (NAMONE e TIMBANE, 2017).

Outra questão que justifica o fraco crescimento da variedade guineense do português é que o crioulo é ainda a língua do dia a dia e responde plenamente às necessidades comunicativas dos guineenses, inclusive nas escolas, em que tanto os alunos como os professores usam-na para se comunicarem, embora não seja a língua oficial. Dessa forma, tanto o contexto histórico de dominação, o qual pode ter criado uma relação não identitária para os nativos, já que o português era a língua usada pelos dominadores, quanto o uso prático do crioulo, atendendo a todas as necessidades cotidianas, não permitiram que o PTL se arraigasse na fala nativa, fazendo com que aspectos mais específicos, periféricos e menos usuais não fossem apropriados pelos guineenses, como o mecanismo da concordância nominal de número.

No que se refere ao segundo aspecto e à sua relação com a apropriação dos aspectos gramaticais do português, os autores destacam que o PGB recebe influências das línguas bantu e do crioulo falado pela maioria da população. Essas interferências influenciam a fonética e a fonologia, as questões semântico-lexicais e as morfossintáticas, incluindo aí a concordância nominal de número.

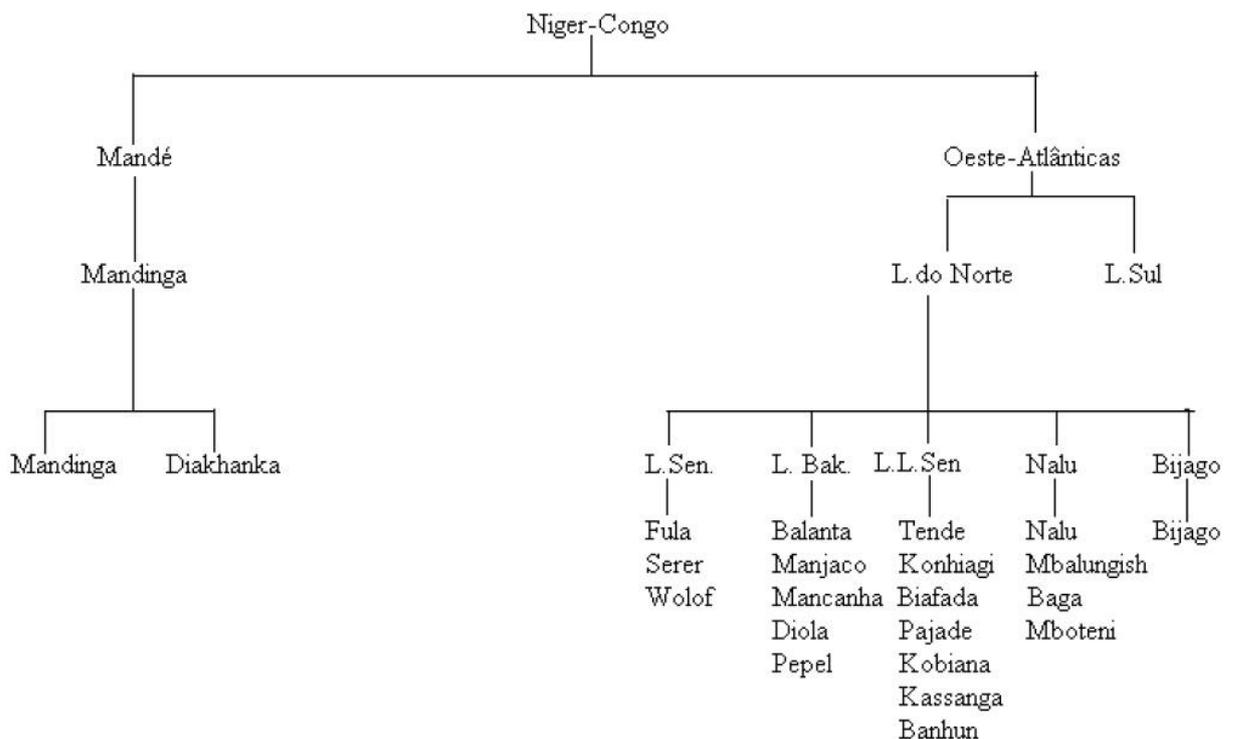
O multilinguismo, circunstância observada tanto em Guiné Bissau quanto em Timor-Leste, é caracterizado pela existência e pelo uso simultâneo de diversas línguas, bem como se suas variedades, em determinada localidade. Essa característica possibilita que falantes interajam em diversas línguas, demonstrando habilidade de produzir sentenças compreensíveis em várias línguas nativas a depender do seu interlocutor.

Em Guiné Bissau, por exemplo, a situação linguística mostra-se bastante complexa, haja vista que, além do crioulo guineense e do PGB há, aproximadamente, 20 outras línguas étnicas faladas no país (CASTRO, 2013, p. 21). A autora afirma que

A maioria dos guineenses nasce em comunidades em que a língua materna é uma língua africana e a sua aquisição acontece no meio familiar, do contato entre pais e filhos. É por meio das línguas africanas que a tradição e a noção de comunidade são transmitidas e mantidas. Apesar de sua importância para a cultura, essas línguas são desprestigiadas e não há um incentivo de seu uso por parte do governo. (CASTRO, 2013, p. 22)

Conforme Mane (2001, p. 40, apud CASTRO 2013, p. 22) as línguas guineenses pertencem à subfamília⁷ Niger-Congo, a qual se divide em dois grupos: Mandé e Oeste-Atlânticas (ou senegalo-guineenses), conforme se vê na figura abaixo:

Figura 1: as línguas⁸ da subfamília Niger-Congo



Fonte: Mane (2001, p. 42, apud CASTRO, 2013, p. 22).

No contexto plurilíngue da Guiné Bissau, vemos o PGB como língua oficial, mas não como língua “oficialmente” adotada pelos guineenses. Isto é, o PGB não constitui uma língua nacional, sendo essa posição ocupada pelo crioulo guineense,

⁷ Subfamílias podem ser encontradas em uma mesma família linguística, isto é, em um mesmo grupo de línguas que compartilham a mesma língua-mãe. Suas semelhanças entre si e com a língua-mãe podem variar.

⁸ Legendas de Mane (2001, p. 42, apud CASTRO, 2013, p. 22-23): L.Sul = línguas do sul; L. Sen. = línguas senegambianas; L.L. Sen. = línguas de Leste ou Senegal Guiné; L. Bak = línguas Bak.

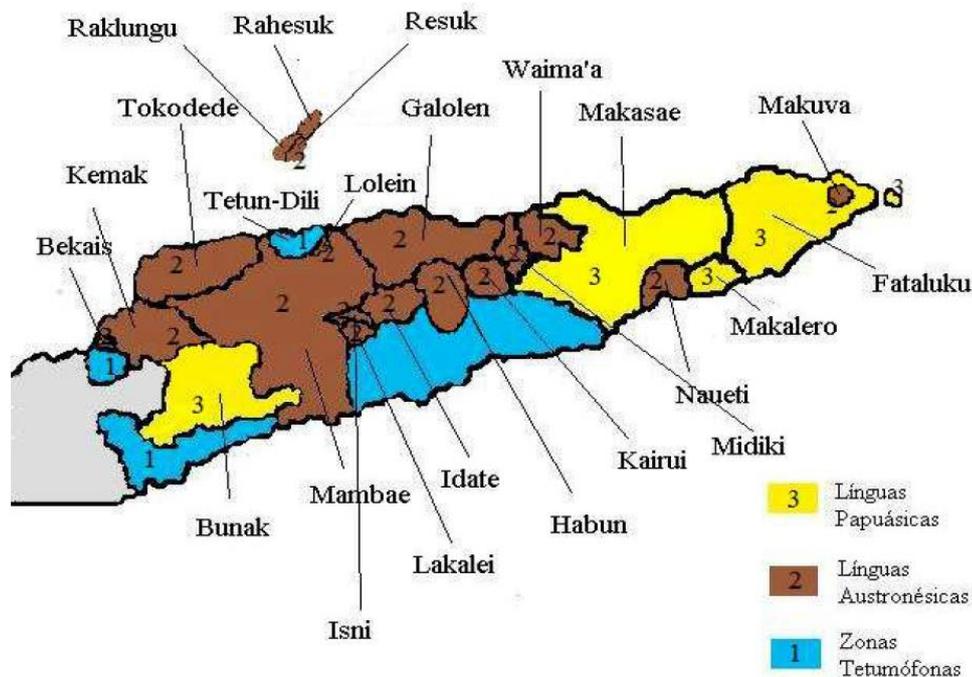
uma vez que se trata da língua pelos diversos grupos étnicos para interação (CASTRO, 2013, p. 24). Estudos de Couto (1994, apud CASTRO, 2013, p. 24) indicam a seguinte distribuição para o uso das línguas na Guiné Bissau: o crioulo guineense é falado por cerca de 75% da população, enquanto outras línguas étnicas são utilizadas por 23% dos guineenses; o PGB, por sua vez, é falado por menos de 13% da população do país.

2.3.1.2 O caso específico do Timor-Leste

Em relação ao Timor-Leste, a situação de multilinguismo também é evidenciada pela existência de línguas com mais de uma filiação genética, a saber: austronésia e papuásica, as quais originaram uma série de outras famílias linguísticas e subfamílias.

O mapa a seguir apresenta a distribuição das línguas de Timor-Leste a partir de suas filiações genéticas:

Figura 2: Timor-Leste e a distribuição das línguas nativas pelo seu território



Fonte: Albuquerque (2011, p. 44).

O multilinguismo vivenciado pelos indivíduos timorenses inicia-se no ambiente familiar. Albuquerque (2011, p. 45), nos diz que “partindo do ambiente familiar de

aquisição de língua pela criança leste-timorense, deve-se distinguir certos conceitos culturais do que acontece realmente no dia a dia do cidadão de Timor-Leste, ou seja, a teoria da prática”. Pelo fato da cultura imaterial do Timor-Leste estar amparada em valores patriarcais (com supremacia do homem, da figura paterna e do filho primogênito, por exemplo), esperar-se-ia que, na cultura do povo leste-timorense, a língua que deveria ser adquirida pelos filhos de um casal fosse, sempre, a língua do pai (ALBUQUERQUE, *Ibid*). No entanto, a escolha por transmitir a língua paterna nem sempre é predominante, uma vez que diversas situações podem provocar o uso de diferentes pelo pai e pela mãe para comunicação com a criança.

Sobre esse assunto, Romaine (1995, p.183, apud ALBUQUERQUE, 2011, p. 45-46) lista seis fatores que podem influenciar na escolha da língua que será utilizada para comunicação com a criança leste-timorenses, os quais são apresentados a seguir:

- Uma pessoa / uma língua: situação em que os pais falam diferentes L1, mas possuem certa fluência na língua do outro. Nessa situação, a língua comunitária é variedade de um dos pais, e cada um deles usa sua própria língua para se comunicar com a criança.
- Língua domiciliar não dominante / uma língua – um ambiente: mesma situação apresentada no item anterior, com a diferença da língua usada não ser a língua comunitária.
- Língua domiciliar não dominante sem apoio comunitário: quando os pais falam a mesma língua, mas esta não é a língua comunitária.
- Língua domiciliar não dominante dupla sem apoio comunitário: situação em que os pais falam línguas diferentes e nenhuma das duas é a língua comunitária.
- Pais não nativos: os pais falam, ambos, a mesma língua, que é a comunitária, mas um deles interage com criança usando uma terceira língua, distinta.

- Línguas mistas: quando tanto os pais quanto a comunidade são bilíngues, e cada um deles utilizam diferentes línguas para se comunicarem com a criança.

(ROMAINE, 1995, p. 183, apud ALBUQUERQUE, 2011, p. 45-46)

Embora haja tantas possibilidades de aquisição de língua pelas crianças leste-timorenses, Albuquerque (2011, p. 46) afirma que a língua tétum encontra-se presente em toda a sociedade de Timor-Leste, podendo ser, nos variados contextos de aquisição no país, adquirido como L1 ou L2 em diferentes situações de diglossia. Esse contexto coloca o PTL numa situação em que este pode ser adquirido não como L2, mas como língua estrangeira, a depender da idade de sua aquisição.

Um resumo do estatuto das línguas em Timor-Leste é feito por Albuquerque (2011, p. 47): o autor apoia-se na constituição de 2002 da República Democrática de Timor-Leste para afirmar que a língua portuguesa e o tétum constituem línguas oficiais do país, enquanto a língua inglesa figura como uma língua de trabalho. O autor afirma, ainda, que

o tétum prasa é falado como L1 por 18% da população, enquanto 82% da população leste-timorense possui alguma fluência nele, pois é a língua que funciona como língua franca, ou seja, a língua de comunicação entre os diferentes grupos etnolinguísticos que possuem L1 distintas. (ALBUQUERQUE, 2011, p. 47)

Em adição, é afirmado pelo autor que a língua portuguesa e a língua inglesa não possuem falantes nativos, sendo a primeira falada por uma pequena parcela da população⁹, de 5%, e, a segunda, de 2% (ALBUQUERQUE, 2011, p. 48).

No contexto multilíngue de Timor-Leste, há diversas línguas nativas, que são L1 de grupos de falantes e possuem um estatuto domiciliar/familiar; há o tétum, que possui o alo de língua nacional/ interétnica/ língua franca; e há as línguas estrangeiras (como o português e o inglês), as quais são consideradas línguas

⁹ O autor salienta que a parcela da população leste-timorense falante da língua portuguesa e da língua inglesa vem crescendo significativamente nos últimos anos.

internacionais/ de cultura e gozam de alto prestígio social (ALBUQUERQUE, 2011, p. 48).

2.3.2 Aquisição *versus* aprendizagem

Todos os seres humanos, ao nascerem, têm a capacidade inata para adquirir uma ou mais línguas graças a um Dispositivo de Aquisição da Linguagem (também denominado de Gramática Universal) que se encontra na Faculdade da Linguagem (FL) localizada como um dos módulos da mente-cérebro desses falantes, de acordo com Chomsky (1986). O modelo da Gramática Universal (GU) considera que há a existência de um período crítico para aquisição de linguagem e que, após a puberdade, a capacidade de aquisição de uma segunda língua diminui.

Sobre esse assunto, Santos (2013, p. 41) afirma o seguinte:

Segundo a perspectiva inatista [...], a aquisição da linguagem ocorre ainda nos primeiros anos de vida do falante em virtude da existência de um “órgão mental” responsável pela linguagem, a saber: a Faculdade da Linguagem. Nesse sentido, o ser humano já nasce com um dispositivo genético para adquirir uma língua presente nessa faculdade, a que denominamos Gramática Universal. (SANTOS, 2013, p. 41)

Com base nessas reflexões, acreditamos que, passado o período crítico, já não é mais possível a aquisição de uma L1 (ou de uma L1 e uma L2 simultaneamente), uma vez que se trata de um processo natural. Dessa forma, após o período crítico, a língua, não mais seria adquirida naturalmente, mas, aprendida, de forma consciente, e, portanto, colocada na posição de língua estrangeira (LE).

O “fator idade” é, frequentemente, uma das variáveis mais discutidas quando se pretende explicar as diferenças que rodeiam a aquisição de língua em contextos multilíngues. Isso porque, para a maioria dos falantes nativos de uma determinada língua, a aquisição acontece a partir do nascimento, enquanto o momento da aprendizagem de uma segunda língua pode variar bastante de indivíduo para indivíduo. Essa informação corrobora com a existência de um “período crítico”, o qual se prolonga da primeira infância do indivíduo até a sua puberdade, momento da vida humana em que o cérebro começa a perder sua elasticidade e, portanto, a capacidade de adquirir, naturalmente, uma língua.

O PGB e o PTL podem ser caracterizados, portanto, como uma segunda língua e uma língua estrangeira nos contextos dos países aqui estudados, respectivamente.

2.4 ESTUDOS DA CONCORDÂNCIA NOMINAL DE NÚMERO EM VARIEDADES DO PORTUGUÊS

Diversos estudos no âmbito da concordância nominal têm sido realizados ao longo dos anos, sobretudo, no que diz respeito ao PB e ao PE. A importância de estudos que tratam de aspectos morfossintáticos de variedades do português mostra-se na possibilidade de compreender os processos que provocam, conforme dito por Brandão (2016), uma “polarização sociolinguística”, a qual, por sua vez, contribui para a formação identitária de países lusófonos da África cuja independência da metrópole portuguesa é ainda recente.

2.4.1 Nas variedades europeia e brasileira

O português europeu, “primeira versão” de uma língua que se disseminaria por outros continentes através de processos de expansão territorial e colonização, perdeu muito de suas características à medida que foi entrando em contato com outras línguas nativas dos países colonizados. O estudo de Bacelar do Nascimento et al. (2008, p. 379-383) nos mostra, por exemplo, desvios gramaticais em variedades africanas do português que não seriam observados no PE. Tais desvios podem ser, segundo a autora, referentes ao léxico e à sintaxe (simultaneamente), somente à sintaxe ou, ainda, à morfossintaxe, como vemos nos exemplos de 8 a 14 a seguir¹⁰:

I - Léxico-sintaxe:

(8) “*para dar os filhos de comer*” G (O)

¹⁰ Nos exemplos de desvios gramaticais apresentados nos exemplos de 8 a 14, (O) refere-se a dados de oralidade; “G” refere-se à produção de falante da Guiné-Bissau; “A”, de falante da Angola; “M”, como já visto nos exemplos de 1 a 7, à produção de falante de Moçambique; “CV”, de falante de Cabo Verde; e “ST” refere-se à produção de falante de São Tomé e Príncipe.

(9) “*querem bater as pessoas*” A (O)

(10) “*pedir uma tia nossa para tomar conta de nós*” M (O)

II - Sintaxe:

(11) “*se faz o tratamento da lepra ali*” G (O)

(12) “*mas isso, se verifica sobretudo [...], nas camadas mais pobres*” CV (O)

III - Morfossintaxe:

(13) “*tu passas e vê o lixo*” G (O)

(14) “*ocê quer receber teu dinheiro*” ST (O)

(BACELAR DO NASCIMENTO ET AL, 2008, p. 379-383)

Para efeitos deste estudo, centraremos a nossa atenção no que diz respeito à morfossintaxe e à variação que pode ser observada no sintagma nominal quanto à concordância de número. Em relação à concordância de número, há diversos modos para se expressar essa categoria no sintagma nominal (doravante SN), entretanto, segundo Vieira e Brandão (2014, p. 88), apenas aquele que pressupõe a colocação do morfema de número em todos os constituintes flexionáveis do SN é que caracteriza o padrão de prestígio, como visto em 15¹¹:

(15) Maria se dá bem com [todass aquelass suass vizinhass estrangeirass].

(VIEIRA e BRANDÃO, 2014, p. 88)

¹¹ Neste texto, Vieira e Brandão (2014, p. 88) não informam de onde extraíram o exemplo (8). Acreditamos ser um exemplo criado pelas próprias autoras para demonstrarem padrões que podem ser verificados, isto é, que são possíveis, no PB e no PE.

No exemplo mencionado, verifica-se que todos os termos flexionáveis recebem o morfema de pluralidade, mas há, ainda, a possibilidade (às vezes, menos prestigiada) dessa marcação acontecer de maneira parcial, quando o morfema de número /s/ apenas é aplicado em um dos constituintes que poderiam receber essa flexão, conforme vemos nos exemplo de 16 a 19¹²:

(16) as menina

(17) as pessoa branca

(18) os meu filho

(19) esses transporte público

Os estudos de Brandão (2013, apud VIEIRA E BRANDÃO, 2014, p. 88) nos permitem concluir que todas essas formas padronizadas nos exemplos de 9 a 12, ainda que haja variação em sua frequência e produtividade, são possíveis no PB. Em relação ao PE, por sua vez, a autora afirma que o padrão se comporta como em 8, com violações raramente registradas, sendo a regra, portanto, categórica.

Como já vimos, a frequência da produtividade de casos de variação é classificada, em Labov (2003, p. 243), como categórica, semicategórica e variável, sendo 100% para regras categóricas, não havendo violações na fala natural; 95 a 99% para semicategóricas, com violações raras, mas possíveis de serem relatadas; e 5 a 95% para regras variáveis. Vieira e Brandão (2014, p. 85-86) questionam a atribuição de dois tipos de aplicação de regras com um intervalo tão arbitrário, de até 5% para regras variáveis, e defendem que, em abordagens qualitativas, é possível tratar aplicações semicategóricas como categóricas, sobretudo quando se trata do PE, em que a utilização de marcas de plural tem seu uso generalizado. No PB, todavia, a regra aplica-se, geralmente, de maneira variável. Apenas quando são inseridos critérios como nível de escolaridade é que os dados obtidos pelas autoras revelam que a regra mostra-se semicategórica, com uma produtividade de 97,2 % entre informantes com Nível Superior.

¹² Os exemplos de (16) a (19) são autorais, a exemplo do que vimos em Vieira e Brandão (2014, p. 88).

Quanto ao PE, a investigação realizada por Vieira e Brandão (2014, p. 92) revela que, num total de quase sete mil sintagmas nominais produzidos por 54 indivíduos, houve apenas oito casos de não utilização da marcação de número em constituintes flexionáveis, considerando todos os critérios de separação dos grupos (faixa etária, escolaridade, etc), o que aponta ser a concordância nominal de número com marcação de pluralidade em todos os termos flexionáveis uma regra categórica, se analisarmos qualitativamente. As autoras afirmam, ainda, que, a partir do que Labov (2003, p. 243) assinala como regra categórica, é possível caracterizar casos prototípicos de marcação variável como hesitação ou lapsos, por exemplo.

Em relação ao PB, um trabalho precursor sobre esta variedade foi realizado por Scherre (1988). Em sua pesquisa, a autora desenvolve análises de variáveis linguísticas baseadas em duas perspectivas: a de análise atomística, que considera cada elemento do sintagma nominal como um dado de análise; e a de análise não atomística, que considera o sintagma nominal inteiro como única unidade de análise. Segundo a autora,

as duas perspectivas de análise a serem apresentadas têm como objetivo mostrar que (1) algumas variáveis atuam especificamente sobre cada elemento do SN; (2) outras atuam principalmente sobre o SN como um todo e (3) um terceiro tipo atua tanto sobre os elementos do SN quanto sobre o SN inteiro. (SCHERRE, 1988, p. 61)

Em seu estudo, Scherre (1988, p. 63) trabalhou as seguintes variáveis linguísticas:

1) Processos morfofonológicos de formação do plural; 2) Tonicidade dos itens lexicais singulares; 3) Número de sílabas dos itens lexicais singulares; 4) Posição linear do elemento no SN; 5) Classe gramatical do elemento nominal; 6) Marcas precedentes ao elemento nominal analisado; 7) Contexto fonético/fonológico seguinte ao elemento nominal sob análise; 8) Função sintática do SN (codificada em cada um de seus constituintes); 9) Animacidade dos substantivos; 10) Grau dos substantivos e dos adjetivos e 11) Formalidade dos substantivos e dos adjetivos. (SCHERRE, 1988, p. 63)

Os resultados obtidos por Scherre (1988) acerca do PB revelam que fatores como a linearidade, por exemplo, têm papel fundamental na realização da

concordância plural. Em outras palavras, segundo a autora, o sintagma nominal que determina as flexões número-pessoais do verbo é aquele que se encontrar mais próximo a ele na cadeia linear da sentença (SARAIVA; BITTENCOURT, 1987, p. 21, apud SCHERRE, 1988, p. 415). Portanto, assumimos em Scherre (1988) que há evidências de que princípios ou motivações externas e internas à língua atuam sobre o seu uso e favorecem determinadas ocorrências, como as de concordância observadas no âmbito do sintagma nominal.

2.4.2 Nas variedades africanas

Conforme já mencionamos, diversos estudos detêm sua atenção ao comparar o PB com o PE, e poucas pesquisas investigam outras variedades do português. Devido à situação de contato linguístico com outras línguas naturais, as variedades africanas do português possuem características peculiares, entre elas, o fato de o português ser a língua oficial e ser usada, em geral, na administração e no ensino, sendo adquirido por muitos como segunda língua e em um modelo que recebe influência do PE (PETTER, 2009, p. 203). De acordo com a autora, esse modelo prevê que as línguas que participam do contato estão numa situação assimétrica: há uma língua matriz, que fornece a estrutura morfossintática e uma(s) língua(s) encaixada(s), que participa(m) com o léxico, principalmente (PETTER, 2009, p.206).

Conforme afirma Petter (2009, p. 207), em português, o determinante dependeria do núcleo nominal para a informação de número, enquanto, em algumas línguas nativas africanas, o determinante seria um prefixo associado à raiz nominal que dependeria do núcleo nominal para formar a classe da palavra. Essa diferença sutil, em contato com variedades africanas do português, pode acrescentar características e formas de concordância nominal não encontradas no PB e no PE.

Em relação à concordância de número, os morfemas de conteúdo constituídos por nomes influenciam a relação entre determinante e núcleo do sintagma nominal. Essa mesma característica, consoante Petter (2009, p. 207), pode ser observada em variedades africanas do português.

2.4.2.1 Angola e Moçambique

O português falado em países da África e em Timor Leste, no continente asiático, constitui, assim como o PB, segundo Petter (2009), “uma língua não nativa, transplantada”, que é resultado de um processo de colonização e que está em constante contato com outras línguas, sejam elas línguas crioulas, indígenas ou outras. Estudos contrastivos entre variedades do português contribuem para explicar como ocorrem variações em padrões morfossintáticos na variedade de cada país, assim como eventuais processos de mudança.

No que tange à concordância nominal, fenômeno em investigação neste trabalho, o português angolano (PA) e o português moçambicano (PM) demonstram características semelhantes ao PB, uma vez que a concordância do núcleo do sintagma nominal com o seu determinante pode variar. Os *corpora* investigados por Petter (2009) mostram diversos casos em que a concordância de número no sintagma nominal não é atestada, conforme podemos ver em 20 e 21:

(20) Você chega lá **os caminhão todo** um dia tão abastecer (Ze53-18/238, apud CHAVAGNE, 2005, p. 240). **(PA)**

(21) Era ele com os outro que descarregava **os tambor** de óleo (LABAN, 1999, p. 141). **(PM)**

(PETTER, 2009, p. 207)

De acordo com Petter (2009), ao observarmos a estrutura do quimbundo, língua do grupo banto falada em Angola, podemos estabelecer modelos para analisarmos as demais línguas faladas por esse grupo. A autora observa que a estrutura do sintagma nominal no quimbundo é composta por um prefixo de classe seguido de uma raiz nominal, como vemos a seguir:

(22) **mù-tù**
(1) *pessoa*

(23) **à-tù**
(2) *pessoas- 'pessoas'*

(PETTER, 2009, p. 209)

Nos exemplos acima, os números (1) e (2) entre parênteses indicam a classe a que pertence o nome, sendo (1) referente a humano – no singular – e 2 referente a humanos, no plural. Conforme vemos, o prefixo à esquerda do nome (que está em negrito) é que indica o número, singular ou plural, ou seja, não há um morfema específico inserido no nome que indique pluralidade, como ocorre no PB com o morfema /s/. Por isso, quando observamos a ausência de concordância nos exemplos (22) e (23) no PA, verificamos que a marca de plural só ocorre no primeiro elemento do sintagma nominal.

2.4.2.2 Guiné Bissau

Assim como ocorre com outras variedades do português, o português de Guiné Bissau (PGB) possui suas particularidades devido ao contato com o crioulo guineense, além de outras línguas existentes no país. Como não há uma quantidade significativa de estudos com dados escritos do PGB, até onde temos verificado, nesta seção, analisaremos a organização do sintagma nominal do crioulo guineense, a fim de compreendermos, posteriormente, a partir da análise dos *corpora* desta dissertação, os padrões de concordância nominal de número encontrados e o tipo de regra linguística (se é categórica, semicategórica ou variável).

Em relação à constituição do crioulo guineense, Castro (2013, p. 24) nos diz o seguinte:

O CG, mais conhecido como *kriyol*, é um crioulo de base portuguesa que surgiu do contato do português com as línguas étnicas guineenses. Junto com o crioulo caboverdiano e o crioulo de Casamansa forma o Grupo Crioulo da Alta Guiné, considerado o mais antigo grupo de línguas crioulas de base portuguesa. Existem várias hipóteses que procuram explicar a origem do CG. Algumas defendem que ele teve origem em Cabo Verde e depois foi transferido para Guiné- Bissau; outras argumentam que foi em Portugal com a ida de escravos para lá e depois teria ido para a África. Há ainda os estudiosos que alegam que o CG surgiu mesmo em Guiné-Bissau. (CASTRO, 2013, p. 24)

No crioulo guineense, segundo Intumbo (2007), a categoria gramatical do núcleo do sintagma nominal pode ser um nome ou um pronome. É em torno desse núcleo que se encontram os seus determinantes e modificadores, com os quais o núcleo mantém relações de concordância. Em relação ao fenômeno da

concordância nominal de número, os afixos têm grande importância na marcação de pluralidade. Assim, o plural pode ser marcado morfológicamente nos nomes a partir da inserção do sufixo -s, o qual é um morfema indicador de plural, assim como ocorre com o português. Observem-se os exemplos (24) e (25) extraídos de Intumbo (2007, p. 36) em que o nome está no singular e no plural, respectivamente:

(24) mininu **djiru**____
 menino inteligente
 ‘um menino inteligente’

(25) mininu-**s djiru**____
 menino-PL¹³ inteligente:SG¹⁴
 ‘uns meninos inteligentes’

De acordo com o mesmo autor, outra característica relevante do crioulo guineense é que, quando o núcleo nominal possui o traço [+humano] e é precedido de um quantificador, este revela tratar-se de plural, como em 19:

(26) dus mininu (-s) **djiru**____
 dois menino (-PL) inteligente
 ‘dois meninos inteligentes’

(INTUMBO, 2007, p. 36)

Ademais, o autor afirma que determinantes e modificadores mantêm-se invariáveis na concordância de número e não recebem, portanto, nenhuma marcação morfológica de número conforme verificamos nos exemplos acima. Isso caracteriza uma evidência da interferência do crioulo guineense no PGB, visto que, neste, não há variação nos determinantes, enquanto naquele, é obrigatório que o determinante varie.

Outras características do CG foram levantadas por Castro (2013) e algumas delas, das que se referem ao sintagma nominal e às questões de concordância de gênero e número, estão explanadas nos parágrafos seguintes. Em síntese, no SN

¹³ PL = plural.

¹⁴ SG = singular.

do crioulo guineense, segundo a autora, há diversos constituintes, dentre eles: 1) os determinantes; 2) os pronomes demonstrativos; 3) os possessivos; 4) os pronomes indefinidos; 5) os pronomes interrogativos e 6) os adjetivos.

A classe dos determinantes não é utilizada em todos os casos e, quando são, fazem concordância com o nome somente nos casos em que há um modificador, como um adjetivo, conforme os exemplos de (27) a (29), abaixo:

(27) **Un** girafa

ART¹⁵ girafa

=_uma girafa'

(28) Mindjer gorda

mulher gorda.F¹⁶

=_mulher gorda'

(29) Omi gordu

homem gordo.MASC¹⁷

=_homem gordo'

(CASTRO 2013, p. 98)

Em relação aos pronomes demonstrativos, não se flexionam nem em gênero nem em número, como é possível vermos nos exemplos (6) e (7) que seguem:

(30) **kil** mininu-s

DEM¹⁸ menino-PL

=_Aqueles meninos'

(31) **es** omi

DEM homem

=_esse homem'

(CASTRO, 2013, p. 99)

¹⁵ ART = artigo (indefinido).

¹⁶ F = feminino.

¹⁷ MASC = masculino.

¹⁸ DEM = demonstrativo (pronome).

No que diz respeito aos possessivos, ocorrem antes do nome e também são invariáveis:

(32) El i **nha** pape
 3SG¹⁹ COP²⁰ 1SG.POSS²¹ pai
 =_Ele é meu pai

(33) El i **si** pape
 3SG COP 3SG.POSS²² pai
 =_Ele é o pai dele'

(CASTRO, 2013, p. 101)

Os pronomes (indefinidos e interrogativos), por sua vez, ainda segundo Castro (2013), são invariáveis quanto ao gênero e ao número, podendo os indefinidos serem posicionados tanto antes quanto depois do núcleo nominal.

(34) **Nin um** livru
 INDEF²³ livro
 =_nenhum livro'

(35) Nha ropa **tudu** N susa
 1SG.POSS roupa INDEF 1SG²⁴ sujar
 =_Toda minha roupa sujei'

(36) **ke ku** anos no kume aonti?
 O que que 1PL²⁵ 1PL comer ontem
 =_O que que nós comemos ontem?'

¹⁹ 3SG = terceira pessoa do singular.

²⁰ COP = copulativo (verbo).

²¹ 1SG.POSS = primeira pessoa do singular e possessivo.

²² 3SG.POSS = terceira pessoa do singular + possessivo.

²³ INDEF = indefinido.

²⁴ 1SG = primeira pessoa do singular.

²⁵ 1PL = primeira pessoa do plural.

(37) Pedro bin di **nunde**?
 Pedro vir de onde
 =Pedro chegou de onde?'

(CASTRO, 2013, p. 102)

Os adjetivos, por fim, compreendem termos que podem variar em gênero, através da oposição –u/-a (para masculino e feminino, respectivamente), e ocorrem, em geral, após o núcleo nominal do sintagma, conforme vemos nos exemplos (38) e (39) com exceção do exemplo (40), no qual observamos o adjetivo antecedendo o nome:

(38) mindjer bunita
 mulher bonita.F
 _mulher bonita'

(39) omi bunitu
 homem bonito.Masc
 _homem bonito'

(40) un **bon** livru
 ART bom livro
 =um bom livro'

(CASTRO, 2013, p. 105)

Em linhas gerais, conforme já visto em Intumbo (2007), Castro (2013) reitera que o acréscimo do sufixo –s ao nome (substantivo ou adjetivo) terminado em vogal modifica-o para o plural. Nos casos de nomes terminados em consoante, por sua vez, a terminação acrescentada é –is.

2.4.2.3 São Tomé e Príncipe

Assim como nas demais variedades africanas do português, a variedade santomense e do Príncipe vivencia uma realidade de contato linguístico com línguas crioulas, numa situação de multilinguismo, no entanto, diferente no que acontece na Guiné Bissau, por exemplo, em que o PGB não é língua materna da maioria dos falantes, em São Tomé e Príncipe é possível encontrar falantes que não falam a língua nativa, o crioulo desse país, pois o português, além de língua oficial, é língua materna da grande maioria da população. Essa característica do português como L1, assim como ocorre no Brasil, aproxima ambos os países no que diz respeito à variação na aplicação de regras de concordância nominal. Essa semelhança indica que pode haver, também, um quadro de polarização sociolinguística no PST, o que possibilitaria que essa variedade desenvolvesse características próprias a partir de sua utilização frequente. Uma dessas características poderia ser a diversificação – o que forneceria outras possibilidades para se indicar pluralidade, não somente a presença do morfema /s/ nos constituintes do sintagma nominal – nos padrões que indicam pluralidade em um vocábulo.

Ainda em relação à situação lingüística de São Tomé e Príncipe, Brandão (2015) verifica que, no PST, a concordância nominal de número possui diversos padrões para a marcação de plural, dentre eles, a forma padrão em que todos os elementos do sintagma recebem a marca de plural, conforme vemos em (41), e casos em que apenas um dos termos recebe essa marca (cf. (42)):

(41) ...nunca me interessei [nos outros clubes]

(42) ...peguei [as minha() ferramenta() nova ()]

(BRANDÃO, 2015, p.37)

O referido estudo traz dados em números percentuais os quais indicam que, entre falantes com baixo nível de escolaridade, a regra de aplicação da concordância nominal mostra-se variável, e, entre falantes com alto nível de escolaridade, semicategórica. Os dados obtidos pela autora apresentam, em adição, que a utilização da marca de número no PST é motivada tanto por aspectos

linguístico-estruturais, como saliência fônica, posição no sintagma e traço de animacidade (fator semântico), quanto por questões extralinguísticas, que incluem nível de escolaridade, sexo e faixa etária; sendo a posição no sintagma e o nível de escolaridade as motivações mais significativas.

2.4.3 No Timor-Leste

Em relação ao português falado em Timor-Leste, pouco sabemos a respeito de seus aspectos morfossintáticos. Grande parte dos estudos linguísticos apresenta resultados semelhantes entre as variedades africanas do Português. No que se refere à concordância nominal, por exemplo, Albuquerque (2012, p. 04), estudando o português de Timor Leste (PTL), faz a seguinte observação:

Sobre a concordância, na norma padrão da língua portuguesa as marcas gramaticais de gênero (masculino x feminino), de número (singular x plural) e do verbo, que deve ser flexionado de acordo com número, pessoa, tempo, modo e aspecto, são obrigatoriamente colocadas em todos os constituintes ligados a um núcleo e devem seguir suas marcas gramaticais (ex. caso o núcleo seja masculino-plural 'meninos' todos os constituintes ligados a ele devem ser marcados como masculino-plural também 'os meninos estudiosos saíram ontem'), assim é possível observar a variação do uso dessas marcas gramaticais em muitas variedades do português. (ALBUQUERQUE, 2012)

O autor afirma, ainda, que

o PTL apresenta variação na marcação do gênero de acordo com os seguintes princípios [...]: simplicidade, a marcação é realizada em Sintagmas Nominais (SNs) mais simples, entendendo aqui SN como elementos que se organizam para formar uma unidade significativa dentro da oração, mantendo relações de dependência e de ordem entre si e em torno de um núcleo nominal [...]; integração, elementos à esquerda e próximos ao núcleo do SN, ou seja, SNs formados somente pelo núcleo nominal (N) ou por um determinante mais o núcleo (Det. + N) tendem a ser mais marcados; e saliência, a marcação mórfica forte, influenciam a realização do gênero.

Em relação à língua tétum, falada por mais de 80% da população leste-timorense, o sintagma nominal organiza-se da seguinte maneira, segundo Albuquerque (2011, p.152):

[ele] pode ser formado pelos seguintes elementos: o núcleo (N), o marcador de posse (POS), a negação (NEG), advérbios (Adv.), determinantes (Det.), modificadores (Mod.) e complementos (Comp.). O único elemento obrigatório é o núcleo (N), que pode ser assumido pela classe dos substantivos, dos pronomes pessoais e demonstrativos. (ALBUQUERQUE, 2011, p.152)

Tal estrutura de formação do sintagma nominal no tétum evidencia, conforme defende Albuquerque (2011, p.178), total ausência de morfologia flexional, não havendo, portanto, marcas morfológicas que indiquem concordância. Como o próprio autor afirma, o português ganha, em geral, no Timor-Leste, o estatuto de língua estrangeira. Assim, justificamos o fato de os dados do PTL fugirem muito do PE, pela grande influência do tétum e por essa evidência da ausência de morfologia flexional.

Na investigação realizada nesta dissertação, notamos como principal semelhança entre o PGB e o PTL, no que se refere à concordância de número, o fato das línguas maternas dos países dessas variedades não possuírem marcas morfológicas que indiquem tais flexões.

Outras características do tétum foram levantadas por Albuquerque (2011) e algumas delas, das que se referem ao sintagma nominal e às questões de concordância de número, serão explanadas nos parágrafos seguintes.

Albuquerque (2011, p.103) considera a marcação de número como instável na língua tétum, uma vez que sua realização acontece pela reduplicação ou pela posposição do pronome de terceira pessoa plural “*sira*” e pelo –s final da língua portuguesa. O autor afirma, ainda, que o emprego dessas formas flexionais está, também, relacionado ao grau de escolaridade do usuário e tem, em adição, relação com as situações de uso da língua. Essas características, conforme Albuquerque (2011, p. 103), são “herdadas” das línguas de superestrato que deram base ao tétum prasa, como o malaio e o português. Alguns exemplos das marcas morfológicas de plural podem ser vistas nas sentenças 43 e 44 a seguir:

(43) estudante **sira** hola buku-**buku** atu estuda
 estudante PL ganhar livro-RED²⁶ IRR²⁷ estudar
 ‘Os estudantes ganharam livros para estudar.’

²⁶ RED = reduplicação.

²⁷ IRR = modo irrealis. Ver Albuquerque (2011).

(44) *funsonarius nasoes unidas nian too iha timoor ohin*
 funcionários nações unidas POS²⁸ chegar LOC²⁹ Timor hoje
 ‘Os funcionários das Nações Unidas chegaram em Timor hoje.’

(ALBUQUERQUE, 2011, p. 103)

Dando continuidade às características gramaticais da língua tétum, podemos dizer, baseados em Givón (1984, p.51 apud ALBUQUERQUE, 2011, p. 122), que em tal língua a classe dos substantivos, os quais, junto aos pronomes, constituem as únicas classes que podem funcionar como núcleos do sintagma nominal, contém os itens que se referem a conceitos estáveis no tempo, os quais, morfossintaticamente, são modificados por sufixos e outros, como aqueles com função de definição de gênero natural (*mane x feto* e *anan x inan*; para masculino e feminino) e plural (*sira*). A classe dos pronomes, no entanto, não possui morfologia flexional de gênero.

Sobre os substantivos em tétum prasa, Albuquerque (2011) nos diz que

os substantivos em TP possuem um traço semântico de [+ definido] e singular, quando não acompanhado por nenhum elemento. Assim, o quantificador numeral *ida* ‘um’ é gramaticalizado e colocado após os substantivos para expressar principalmente a ‘indefinição’, porém há algumas situações orientadas pelo discurso que *ida* pode marcar ‘singular’.
 (ALBUQUERQUE, 2011, p. 102)

Outros constituintes do SN no tétum prasa são os determinantes, sendo os mais comuns os demonstrativos e os quantificadores, os quais são utilizados logo após o núcleo do SN e indicam, em relação ao nome, características como plural, gênero, indefinição, diminutivo, entre outras.

A partir do que lemos em Albuquerque (2011), apesar da instabilidade a qual o autor se refere no que diz respeito à marcação de plural no tétum, sintetizamos que a pluralização dos termos nessa língua é feita com a reduplicação ou com a posposição do pronome de terceira pessoa plural “*sira*” e pelo –s final da língua portuguesa, conforme vimos nos exemplos 43 e 44.

²⁸ POS = possessivo.

²⁹ LOC = preposição locativa.

3 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa foi empreendida adotando-se como base teórico-metodológica o modelo laboviano. No presente capítulo, descreveremos os processos envolvidos em tal metodologia no que diz respeito à constituição dos *corpora*, aos critérios utilizados para a escolha das amostras e ao método de análise utilizado. Também caracterizaremos o local de coleta dos dados e apresentaremos as variáveis linguísticas e extralinguísticas selecionadas.

3.1 MÉTODO UTILIZADO

Para a realização desta pesquisa, adotamos o método de abordagem o indutivo. Quanto aos métodos de procedimento, foram utilizados o comparativo (comparação entre o PGB e o PTL) e o estatístico (obtenção dos dados quantitativos que serão analisados à luz da Sociolinguística). Portanto, realizamos a análise quantitativa dos dados e o estudo contrastivo entre o PGB e o PTL, tomando por base o fenômeno linguístico da concordância de número no sintagma nominal e a possível interferência da morfossintaxe do crioulo guineense e do tétum sobre essas variedades.

Para tanto, analisamos sintagmas nominais selecionados a partir de redações produzidas por estudantes guineenses e timorenses no âmbito do vestibular do ano de 2013 da Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB), uma vez que tal ano consistia no único em que haviam ingressado estudantes timorenses em número relevante para comporem os dados de análise desta pesquisa. Nos anos seguintes, o contingente de estudantes timorenses a ingressarem na UNILAB diminuiu drasticamente, havendo, inclusive, várias edições do certame em que nenhum candidato oriundo do Timor-Leste tenha ingressado.

É relevante mencionarmos que as propostas temáticas das produções textuais não foram consideradas, uma vez que utilizamos, apenas, as sentenças produzidas pelos escreventes (falantes nativos das variedades do português selecionadas para análise). Trabalhamos, assim, com os *corpora* sincrônico, a fim de compreendermos as características das variedades em questão, verificando se, nos textos produzidos em PGB e PTL é possível encontramos interferências morfossintáticas do crioulo guineense e do tétum, respectivamente.

3.2 LOCAL DA COLETA DOS DADOS

Para a realização do presente estudo, utilizamos as redações contidas no banco de dados da UNILAB³⁰, localizada no estado do Ceará, onde há campi nas cidades de Redenção³¹ (campus da Liberdade, que é o campus-sede da instituição; e campus das Auroras) e Acarape (campus dos Palmares), que ficam a, aproximadamente, 60km de Fortaleza, capital do estado e localizam-se da região do Maciço de Baturité.

Essa universidade começou a ser idealizada no ano de 2008 quando foi criada uma Comissão para sua implementação. A partir dali, empreenderam-se, nos dois anos que sucederam a criação dessa Comissão, estudos acerca de questões comuns ao Brasil e aos demais países lusófonos, bem como fez-se o levantamento de atividades de planejamento e organização acadêmica em geral. No dia 20 de julho de 2010, a UNILAB é criada pela Lei nº 12.289, sancionada pelo então presidente Luís Inácio Lula da Silva, e instalada em 25 de maio de 2011.

Um aspecto relevante acerca da UNILAB é que ela surgiu com o objetivo de capacitar cidadãos que pudessem contribuir com a integração entre o Brasil e os demais países que integram a Comunidade dos Países de Língua Portuguesa (CPLP), a saber: Angola, Cabo Verde, Guiné-Bissau, Moçambique, Portugal, São Tomé e Príncipe e Timor-Leste, promovendo, assim, o intercâmbio entre cultura, ciência e educação de modo a contribuir com o desenvolvimento destes.

Os cursos ofertados pela referida universidade abarcam diversas áreas e contam com um corpo docente e discente composto por brasileiros e estrangeiros, os quais contribuem para que a integração acadêmica possa originar políticas públicas que possibilitem a mitigação das desigualdades sociais e um relevante avanço nas políticas de cooperação e internacionalização do Ensino Superior.

³⁰ Dados obtidos diretamente do site da UNILAB. Disponível em: www.unilab.edu.br. Acesso em: 16 ago. 2018.

³¹ A cidade recebe esse nome por ter sido a primeira cidade brasileira a libertar todos as pessoas escravizadas.

3.3 SUJEITOS DA PESQUISA, AMOSTRA, E CONSTITUIÇÃO DOS CORPORA

No presente estudo, os dados foram obtidos de um banco de dados já existente na UNILAB, mais precisamente, da Coordenação de Políticas de Acesso e Seleção (CASE), que fica na Pró-reitoria de Graduação (PROGRAD). Esse banco continha todas as redações produzidas por candidatos a uma vaga na UNILAB-CE para diversos cursos, a saber: Administração Pública, Agronomia, Antropologia, Bacharelado em Humanidades, Ciências Biológicas – Licenciatura, Ciências da Natureza e Matemática, Engenharia de Energias, Física, História, Letras – Língua Portuguesa, Matemática – Licenciatura, Pedagogia – Licenciatura e Sociologia – Licenciatura). Para fins de separação e coleta dos textos, entendemos como “local” o critério de separação “país de origem”, ou seja, a “comunidade” - que constituiu o “local” onde os dados a serem analisados foram coletados -, foi formada pelos candidatos da Guiné-Bissau e do Timor-Leste, escreventes das redações. Esses candidatos (aqui chamados, também, de “informantes” ou “escreventes”) compuseram o grupo social que interessou a esta pesquisa.

Definidos os sujeitos da pesquisa, partimos para a estratificação de nossa amostra, isto é, para “as dimensões sociais relevantes para a variação, pois elas vão se refletir (...) na constituição das células sociais³²” (COELHO et al, 2015, p. 101). Para tanto, adotamos as seguintes variáveis extralinguísticas: a) país de origem (Guiné-Bissau e Timor-Leste), sexo (homem e mulher) e situação no vestibular (aprovado ou reprovado). O Quadro 2, a seguir, apresenta a distribuição desses informantes:

Quadro 2 - Distribuição dos informantes³³ (total = 400 informantes)

País de origem	Guiné-Bissau				Timor-Leste			
	MA	HA	MR	HR	MA	HA	MR	HR
Sexo e								

³² Assumimos, com Coelho et al (2015), que “célula social” é “um conjunto de indivíduos agrupados pelas mesmas características sociais relevantes para a análise de fenômenos de variação e mudança linguística. Defendemos, com os autores, que tais características sociais incluídas no bojo na análise “não são aleatórias, mas seguem os critérios de estratificação social que têm se mostrado relevantes nos estudos sociolinguísticos” (COELHO et al, 2015, p. 101).

³³ No Quadro 2, “MA” refere-se a mulheres aprovadas; “HA”, a homens aprovados; “MR”, a mulheres reprovadas; e “HR”, a homens reprovados.

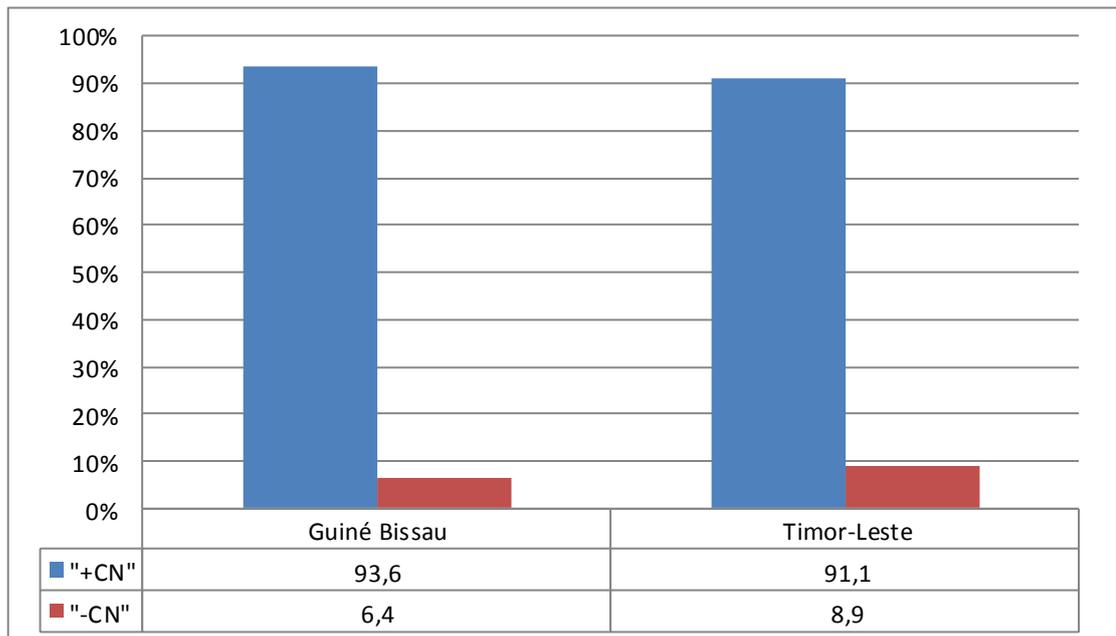
Situação no vestibular	50	50	50	50	50	50	50	50
Total	400							

Fonte: autora desta pesquisa.

Ainda em relação à distribuição de nossos informantes, salientamos que a amostra é aleatória. As redações foram selecionadas somente pelos critérios predeterminados, não havendo leitura prévia de seu conteúdo.

Em relação à primeira língua dos estudantes e à idade com a qual prestaram o exame vestibular, não obtivemos essa informação, pois não foi possível acessar os formulários de inscrição nem aqueles com informações socioeconômicas dos estudantes. Apesar disso, baseamo-nos nas informações já mencionadas neste estudo acerca do quantitativo populacional de falantes do crioulo guineense, na Guiné Bissau, e do tétum prasa, no Timor-Leste, para considerarmos a influência dessas duas línguas nativas no uso da concordância nas variedades do português de cada um desses países (PGB e PTL). Defendemos, portanto, que mesmo não sendo o tétum e o crioulo guineense a primeira língua dos escreventes, essas línguas também foram adquiridas em idade bastante precoce, considerando a sua importância para a população dos países e a identificação dos usuários com elas.

Outro aspecto relevante a ser destacado em relação aos candidatos escreventes, é que eles já passaram pelo último ciclo do que seria o Ensino Médio no Brasil, ou seja, já possuem um maior grau de escolarização. Assim, por terem mais anos de escolarização e já poderem se candidatar a uma vaga na universidade, além de encontrarem-se, no momento da escrita da redação do vestibular, em um contexto monitorado de avaliação, já não são esperadas muitas ausências de concordância de número, e o quantitativo geral da aplicação de concordância tomando por base o país de origem, conforme vemos no Gráfico 1, a seguir:

Gráfico 1 - Percentual de +CN e -CN, tomando por base o país de origem

Fonte: autora desta pesquisa.

Apesar de não termos realizado entrevistas sociolinguísticas nesta pesquisa, alguns aspectos metalinguísticos puderam ser observados pelo fato de os alunos guineenses e timorenses serem levados a escrever sobre o uso da língua portuguesa em seus respectivos países, suscitando neles, por exemplo, discussões sobre como o português se organiza e se distribui em seus países, levando em conta sua importância e aceitação. Se, por um lado, uma proposta de redação que aborda um assunto de natureza metalinguística inclina a atenção do usuário para “como” as coisas estão sendo ditas por ele, o que propicia maior grau de monitoramento ao uso, por outro, um conteúdo constituído pelo relato de experiências pessoais diminuem esse grau de monitoramento. Essa situação também pôde ser observada nas redações analisadas, uma vez que, em algumas delas, os candidatos discorriam sobre sua situação de estar prestando um vestibular e escrevendo em uma língua que não era a sua primeira (não era sua L1). Em geral, esses estudantes compartilham práticas sociais fundamentadas não somente em aspectos étnicos, como valores, crenças e modos de agir, mas também nos modos de usar a língua portuguesa.

Definida a amostra, os *corpora*, por sua vez, foram constituídos dos sintagmas nominais presentes nas redações selecionadas em que há a variante padrão e não-padrão relacionada à concordância nominal de número.

Trabalharemos, neste estudo, apenas com o domínio interno do sintagma, sendo dispensado o domínio predicativo. Vejamos dados do PGB e do PTL que evidenciam essas variantes:

- PGB:

- a) Concordância nominal de número:

(1)a. variante padrão

(58) *As línguas dialéticas* (GHA)

(59) *A complexidade dos grupos étnicos* (GHA)

(60) *Os nossos alunos* (GHR)

b. variante não-padrão

(61) *um papel muito interessantes* (GHR)

(62) *os guineense* (GHR)

(63) *nas escola e vias públicas* (GHR)

- PTL:

- b) Concordância nominal de número:

(1)a. variante padrão

(70) *dos professores portugueses* (TMA)

(71) *com os meus esforços* (TMA)

(72) *as outras pessoas* (TMA)

b. variante não-padrão

(73) *o estrangeiros* (TMR)

(74) *os cidadania* (TMR)

(75) *esses país timorense* (TMR)

Vale referirmos, ainda, que, como não haveria contato direto do pesquisador com os sujeitos que produziram as redações, esta pesquisa dispensou o Termo de Consentimento Livre e Esclarecido (TCLE), uma vez que a autorização para uso de dados secundários já constava na carta de anuência emitida e assinada por pessoa responsável na UNILAB-CE.

Salientamos que informações acerca dos sujeitos são confidenciais, havendo autorização apenas para divulgação dos resultados do estudo em eventos ou publicações científicas. Por isso, não há identificação dos escreventes, com exceção do momento em que os textos estavam sendo selecionados a partir das variáveis extralinguísticas, instituídas a partir dos critérios de inclusão do material na investigação proposta (*país de origem, sexo e situação no vestibular*)

Os dados colhidos durante a realização desta pesquisa ficaram armazenados em *pendrive* adquirido exclusivamente para essa finalidade, e os dados que compõem os *corpora* deste trabalho estão sendo mantidos no computador pessoal e em pastas de arquivo, sob a nossa responsabilidade.

3.4 SELEÇÃO DAS VARIÁVEIS

Concluída a etapa de constituição dos *corpora*, as variáveis linguísticas e extralinguísticas foram selecionadas para realizarmos a análise linguístico-

quantitativa dos dados. Para tanto, seguem as variáveis selecionadas com seus respectivos fatores referentes ao fenômeno da concordância nominal de número.

Apoiamo-nos, nesta pesquisa, em algumas das variáveis linguísticas estudadas por Scherre (1988), a saber: classe gramatical do elemento, posição linear do elemento no sintagma e saliência fônica nas três dimensões analisadas por Scherre (1988): processos morfofonológicos de formação de plural, tonicidade do item lexical singular e número de sílabas do item lexical singular. Não utilizamos todas as variáveis analisadas pela autora devido ao tempo que era escasso para a realização desta pesquisa, que já se apresentava bastante robusta pelo volume de dados obtidos. Em adição, a análise pelo viés das três dimensões da saliência fônica já se apresenta como bastante significativa, tendo em vista o quantitativo de constituintes analisados e as rodadas de cruzamentos entre as três dimensões dessa variável.

Para efeitos deste estudo, realizamos a análise de tipo atomística, a qual considera todos os itens flexionáveis como elemento de investigação. Em trabalhos futuros acerca da concordância no âmbito do sintagma nominal, procederemos uma investigação de base não atomística, considerando o SN inteiro como unidade de análise.

3.4.1 Variáveis dependentes

3.4.1.1 De número

- Aplicação da regra de concordância nominal de número (presença do morfema de plural).
- Não-aplicação da regra de concordância nominal de número (ausência do morfema de plural).

3.4.2 Variáveis linguísticas para análise da concordância nominal de número

3.4.2.1 Posição linear do elemento no sintagma nominal

- Posição 1

(82) **os** seus jovens para fora do país (GHA)

(83) nas zonas urbanas e rurais (GHA)

(84) muitas pessoas (GHA)

- Posição 2

(85) os seus jovens para fora do país (GHA)

(86) todos os países (THA)

(87) os seus conhecimentos (THA)

- Posição 3

(88) os seus jovens para fora do país (GHA)

(89) todos os estudantes (TMA)

(90) as nações vizinhas (TMA)

- Posição 4

(91) Todos os países colonizado (GHA)

(92) Nos seus próprios país (GHR)

(93) Da escolas lusófonas africanos (GHR)

(94) Todas as instituições públicas (GHR)

- **Justificativa para a escolha da variável:**

Em Scherre (1988), é defendido que o local, estritamente linear, nos termos da autora, que o elemento analisado ocupa no sintagma nominal favorece a utilização ou não da marca de concordância. Sobre esse assunto, Scherre (1988, p. 143) defende:

Todos os trabalhos realizados até o presente momento apresentam, com relação a esta variável, uma conclusão uniforme: a primeira posição do SN é a mais marcada, num índice probabilístico nunca inferior a 0,70; e as demais posições evidenciam um índice baixo de marcas, estabelecendo-se assim uma oposição forte com relação ao que ocorre com o primeiro elemento do SN. (SCHERRE, 1988, p. 143)

A autora defende, a partir de um levantamento feito com diversos estudos, que a variável posição atua como sendo a mais importante de todas, uma vez que exerce influência uniforme sobre a regra da concordância de número no âmbito do sintagma nominal Scherre (1988, p. 143).

Segundo Biderman (1968, apud SCHERRE, 1988, p. 147), a relevância dessa variável justifica-se, também, por ser um condicionamento em função do fenômeno da redundância, isto é, a (desnecessária) “repetição” explícita de sinais que não trazem nova informação à já existente. Ainda sobre esse assunto, Ali (1971, p. 279, apud SCHERRE, 1988, p. 147) nos diz:

Considera-se, então, de forma geral, que o mecanismo da concordância não constitui uma necessidade lógica das línguas, pois uma marca formal ou semântica de plural em algum ponto no SN é suficiente para se transmitir a informação desejada. Como se afirma que em Português marca-se preferencialmente a primeira posição, as demais marcas tornam-se desnecessárias. (ALI, 1971, p. 279, apud SCHERRE, 1988, p. 147)

Considerando essas informações, a variável “posição linear” caracteriza-se como indispensável neste estudo.

3.4.2.2 Classe gramatical do elemento nominal

- Substantivo

(94) várias **línguas** étnicas (GHA)

(95) nas **escolas** (GHA)

- Artigo definido

(96) **os** professores timorenses (THA)

(97) **as** ciências (GHA)

- Quantificador (CASTILHO, 2010)

(98) **muitas** crianças (THA)

(99) **vários** desafios (TMA)

(100) **todas** as populações (GMA)

(101) **algumas** escolas da língua portuguesa (THA)

- Preposição (ainda que apresentadas como contração)

(102) **dos** professores portugueses (GHA)

(103) **dos** países africanos (GHA)

(104) **pelos** colonizadores (GHA)

- Adjetivo

(105) dos comerciantes **estrangeiros** (GHR)

(106) dos quadros **técnicos** (GMA)

(107) as instituições **nacionais** (GHR)

- Pronome demonstrativo

(108) **estes** fatores (GMA)

(109) **essas** dificuldades (GHR)

(110) **desses** documentos (THA)

(111) **esses** dois línguas (THR)

- Pronome possessivo

(112) **nossos** parceiros (GHA)

(113) **nossos** próprios dirigentes (GMA)

(114) as **nossas** fronteiras (GMA)

(115) as **suas** formas (GHA)

(102) as **suas** experiências (THA)

(103) os **meus** irmãos (THA)

- **Justificativa para a escolha da variável:**

A classe gramatical também constitui uma variável analisada por Scherre (1988), a qual indica que essa variação de marcas de plural no sintagma nominal pode ser favorecida ou inibida partindo da classificação morfológica do elemento analisado. Por exemplo, Cedergren (1973, apud, SCHERRE, 1988, p. 148), ao analisar o Espanhol do Panamá, constatou que, nos casos do -S morfema de plural, os determinantes inibiam a forma 0(zero) e os substantivos e adjetivos favoreciam-na. A autora observou, ainda, que os adjetivos tinham mais chances de reterem a marca do que os substantivos (CEDERGREN, 1973, p.46, apud SCHERRE, 1988, p. 148).

Tendo sido analisada por Scherre (1988), linguista que compõe o acaboço teórico desta dissertação, julgamos relevante incluir esta variável em nosso estudo.

3.4.2.3 Saliência fônica – Dimensão processos morfofonológicos de formação do plural

- itens terminados em -l (oficiaisis, locaisis)
- itens terminados em -ão (cidadão, condição)
- itens terminados em -r (professorere, setoresres)
- itens terminados em -s (paíss, várioss)
- itens terminados em vogal - plural regular (escolaa, colegaa)
- itens terminados em -m (bomm, algunsns)

3.4.2.4 Saliência fônica – Dimensão tonicidade do elemento

- Oxítona e monossílabo tônico (superior, bom)
- Paroxítona e monossílabo átono (língua, os)
- Proparoxítona (dialética, lusófonos)

3.4.2.5 Saliência fônica – Dimensão número de sílabas dos itens lexicais singulares

- Monossílabo (bom, os, as)
- Dissílabo (setor, rural)
- Mais de duas sílabas (timorense, portugueses)
- **Justificativa para a escolha da variável saliência fônica nas três dimensões analisadas por Scherre (1988):**

O estudo de Scherre (1988) nos aponta que a variável “saliência fônica” atua sobre cada elemento do sintagma nominal. A partir desse fator, determinados itens lexicais teriam mais ou menos chances de serem marcados no plural. A autora evidencia que

a sequência menos marcada é seguramente constituída de elementos nominais regulares paroxítonos. A atuação desta variável explica-se pelo Princípio da Saliência que consiste em estabelecer que formas mais salientes são mais perceptíveis e, portanto, mais marcadas. (SCHERRE, 1988, p. 511)

No caso analisado por Scherre (1988), observaram-se mais marcações nas construções que apresentavam maior diferenciação fônica na relação singular/plural ou que tivessem marca de acento na sílaba que receberia o morfema de plural (SCHERRE, 1988, p. 511).

Embora no estudo de Scherre (1988) o número de sílabas dos itens lexicais singulares não tenha se mostrado como relevante na análise dos dados feita pela autora, optamos por incluir esse fator para verificarmos se a mesma “irrelevância” é observada no PGB e no PTL. Justificamos essa escolha pelo fato de os números percentuais apresentados por Scherre (1988), quanto à ausência/presença de concordância entre os elementos do sintagma nominal, tenham demonstrado certa distância, sendo de 95% para concordância com palavras monossílabas, 64% com palavras dissílabas e 58% com palavras com mais de duas sílabas.

Dadas as evidências demonstradas por Scherre (1988) quanto à relevância desta variável, decidimos, também, por incluí-la em nosso estudo para que pudéssemos observar o comportamento do PGB e do PTL em relação ao aspecto saliência fônica.

3.4.3 Variáveis extralinguísticas para análise da concordância nominal de número

3.4.3.1 País de origem

- Guiné Bissau
- Timor-Leste

- **Justificativa para a escolha da variável:**

A partir da leitura de trabalhos realizados sobre a concordância de número no sintagma nominal no PB (SCHERRE, 1988; SCHERRE; NARO, 1998, 2006) que concluem ser um fenômeno variável nessa variedade e ser influenciado por variáveis linguísticas e extralinguísticas, pretendemos, nesta pesquisa, ampliar a discussão a partir de um estudo contrastivo entre duas variedades não europeias do português, a saber: o PGB e o PTL, cujas pesquisas sociolinguísticas no âmbito da concordância nominal ainda são escassas até onde temos verificado. Nesse cenário, a escolha da variável “país de origem” justifica-se, pois contribuirá com os estudos sociolinguísticos e fortalecerá o campo do contato linguístico, em nosso caso

específico, do contato do português como segunda língua (L2) com a língua materna (L1) dos falantes guineenses e timorenses. Esperamos, acerca dessa variável, encontrar resultados semelhantes para os dois países, considerando o número de falantes dessas variedades do português. Em outras palavras, como o português, apesar de ser língua oficial tanto de Guiné Bissau quanto de Timor-Leste, não é a língua mais falada pelos usuários e não se trata do idioma com o qual a população mais se identifica, levantamos a hipótese de que os resultados para o uso da concordância será parecido.

Tal variável contribui com os estudos morfossintáticos da concordância de número no sintagma nominal, sobretudo, em se tratando de um estudo contrastivo entre essas variedades do português. Em adição, enfatiza-se que o estudo sociolinguístico variacionista no contexto dos países escolhidos pode fornecer suporte empírico para que se combatam ideologias preconceituosas e excludentes justificadas pelas diferenças linguísticas.

3.4.3.2 Sexo

- Feminino
- Masculino

- **Justificativa para a escolha da variável:**

Em pesquisas sociolinguísticas, é relevante pensarmos sobre os limites e as formas em que determinados fenômenos estão relacionados ao fator “sexo” dos informantes. Diversos estudos sociolinguísticos apontam que determinadas formas de prestígio, por exemplo, são mais comuns em falas de mulheres, o que sugere um maior grau de consciência acerca do *estatuto* social de determinadas formas linguísticas. Labov (1966a, apud LABOV, 2008, p. 281), nos diz que “na fala monitorada, as mulheres usam menos formas estigmatizadas do que os homens e são mais sensíveis do que os homens ao padrão de prestígio”. Em adição, o autor afirma que “fica claro que as mulheres são mais sensíveis do que os homens aos

valores sociolinguísticos explícitos” (LABOV, 2008, p. 282), daí a relevância de incluirmos a variável sexo em nossa análise.

O papel desempenhado pela mulher numa dada sociedade e o espaço ocupado por ela na organização social revelam padrões de variação que podem diferir de uma comunidade para outra, caso a organização social seja diferente. Paiva (2013, apud MOLLICA e BRAGA, 2013, p. 35), exemplifica essa questão ao citar o conservadorismo linguístico das mulheres emergindo da análise de variações em comunidades de fala ocidentais, em contraposição ao que ocorreria em comunidades de fala marcadas por outros valores sócio-culturais, nos quais a mulher tivesse menos espaço, por exemplo, ou menor representatividade e voz. Segundo essa autora,

Quando se trata de implementar na língua uma forma socialmente prestigiada (...) as mulheres tendem a assumir a liderança da mudança. Ao contrário, quando se trata de implementar uma forma socialmente desprestigiada, as mulheres assumem uma atitude mais conservadora e os homens tomam a liderança no processo. (PAIVA, 2013, apud MOLLICA e BRAGA, 2013, p. 36)

Paiva (2013, apud MOLLICA e BRAGA, 2013, p. 35) defende, ainda que, mesmo “que os padrões de correlação possam diferir, eles refletem mais do que diferenças biológicas, diferenças no processo de socialização e nos papéis que cada comunidade atribui a homens e mulheres”. Metodologicamente, a explicação quanto a esses padrões deve ser feita com bastante cuidado, visto que eles podem estar diretamente associados à dinâmica da comunidade de fala, isto é, a como se estabelece o processo de socialização de homens e mulheres. Em locais onde as diferenças sociais entre homens e mulheres não sejam tão discrepantes, por exemplo, a tendência é que haja menos diferenças na utilização de determinadas formas linguísticas variantes. Já em regiões onde a separação dos papéis desempenhados por homens e mulheres provoca um distanciamento maior entre eles, é possível que as divergências no uso de algumas formas variantes sejam mais evidente. Nesse cenário, assumimos, com Paiva (2013, apud MOLLICA e BRAGA, 2013, p. 41) que “qualquer explicação das diferenças linguísticas entre homens e mulheres deve ser relativizada em função do grupo social considerado”.

A respeito dessa variável, houve dificuldade em coletar redações de estudantes do sexo feminino tanto de Guiné Bissau quanto de Timor-Leste. Isso porque o número de mulheres que são aprovadas no vestibular da UNILAB-CE é bastante reduzido. Na verdade, constatamos que o número é reduzido desde a submissão ao processo seletivo da UNILAB-CE, uma vez que buscamos as redações de candidatas reprovadas também, o que foi igualmente difícil, evidenciando que a variável “aprovação” no vestibular não foi a responsável pela nossa dificuldade na ocasião da coleta, uma vez que o número de mulheres é reduzido desde a decisão de se inscreverem no vestibular. Se houvesse um número expressivo de mulheres reprovadas, saberíamos que a dificuldade estaria em alcançar a aprovação, o que não foi o caso, pois, mesmo entre os estudantes reprovados no exame, havia um número menor de redações feitas por mulheres, em comparação a dos estudantes do sexo masculino.

3.4.3.3 Situação no vestibular

- Aprovado
- Reprovado
- **Justificativa para a escolha da variável:**

A utilização do critério “aprovação” ou “reprovação” em um processo seletivo caracteriza uma variável incomum nos estudos sociolinguísticos, os quais, frequentemente, consideram o nível de escolaridade como uma variável social. Assim, a seleção da variável apresentada neste item será justificada por uma comparação/equiparação com a variável nível de escolaridade, por acreditarmos que elas se aproximam, uma vez que o resultado obtido pelos estudantes guineenses e timorenses no vestibular da UNILAB-CE nos remete à qualidade na escolarização do candidato.

Diversos estudos sociolinguísticos apontam a variável nível de escolaridade como sendo extremamente relevante para a ocorrência e a produtividade dos padrões de variação (GONÇALVES, 2012, apud SEDRINS e SILVA, 2017; BRANDÃO E VIEIRA, 2012). Isso evidencia que quanto mais escolarizado é o falante, maiores serão os índices de aplicação da regra de variação.

Tal variável possibilita, ainda, que a associemos a outras variáveis sociais, quando estas se aplicam, como nível socioeconômico do escrevente, se ele tem ou não acesso à cultura, e, ainda, se ele mantém contato com indivíduos pertencentes a outros estratos sociais, entre outras associações possíveis. Brandão e Vieira (2012) constatam que quanto às variáveis relevantes à compreensão do funcionamento dos dados, destaca-se a atuação da variável extralinguística **nível de escolaridade**, cuja descrição permitirá a comparação com os resultados relativos à concordância nominal.

3.5 TRATAMENTO QUANTITATIVO DOS DADOS

Selecionadas as variáveis, codificamos os dados para tratamento quantitativo através de sua rodada no programa computacional *GoldVarb-X* (SANKOFF; TAGLIAMONTE; SMITH, 2005), através do qual foi possível a obtenção dos resultados estatísticos e probabilísticos para realização da análise linguístico-comparativa.

Obtidos os resultados quantitativos, foi feita a análise linguístico-comparativa dos dados do PTL e do PGB sobre a concordância nominal de número nos dados de escrita, através da qual foi possível percebermos semelhanças e/ou diferenças entre essas duas variedades não-europeias do português, tomando por base os fatores (linguísticos e/ou extralinguísticos) que condicionam o uso da variante relacionada a essa concordância.

Entre as ferramentas que o programa computacional mencionado oferece, está a obtenção dos pesos relativos da ocorrência de variação a partir das variáveis utilizadas. Esse dado varia entre 0,0 e 1,0 e é considerado significativo quando se apresenta maior ou igual a 0,5, indicando que a variável em questão apresenta-se como relevante no caso de variação pesquisado.

4 OS DADOS

Esta investigação propõe-se a realizar um estudo contrastivo entre as variedades guineense e timorense do português, porém, antes de emprendermos a análise contrastiva, no presente capítulo, explanaremos, inicialmente, alguns resultados os quais levam em consideração o quantitativo global dos dados. Neles, apresentamos os resultados obtidos a partir das informações acerca da concordância nominal de número em Guiné Bissau e em Timor-Leste simultaneamente.

4.1 RESULTADOS GLOBAIS

Como já dito no capítulo anterior, este estudo contou com uma amostra de 400 redações de vestibular, sendo 200 produzidas por estudantes guineenses e 200 produzidas por estudantes timorenses, todas escritas em português, mais precisamente, nas variedades do português de cada país (PGB e PTL). Dessas redações, extraímos todos os sintagmas nominais em que houvesse, ao menos, um elemento pluralizado, seguindo a metodologia aplicada por Scherre (1988). Catalogados esses sintagmas e separados os seus constituintes para análise, obtivemos um total de 6506 constituintes analisáveis, cuja distribuição do quantitativo geral e percentual por país e sexo, tomando por base a presença ou ausência de concordância, vemos no Quadro 3, a seguir:

Quadro 3 - Quantitativo geral e percentual de constituintes com “+CN” e “-CN” por país e sexo, obtido do programa computacional *goldvarb-x*

	Guiné Bissau		Timor-Leste		Total	
	Quantitativo	%	Quantitativo	%	Quantitativo	%
Homens	+CN = 1757	94%	+CN = 1222	92%	+CN = 2979	93%
	-CN = 107	6%	-CN = 112	8%	-CN = 219	7%
Total	1864		1334		3198	

Homens						
Mulheres	Quantitativo	%	Quantitativo	%	Quantitativo	%
	+CN = 1755	93%	+CN = 1291	91%	+CN = 3046	92%
	-CN = 132	7%	-CN = 130	9%	-CN = 262	8%
Total Mulheres	1887		1421		3308	
Total Geral	+CN = 3512	94%	+CN = 2513	91%	+CN = 6025	93%
	-CN = 239	6%	-CN = 242	9%	-CN = 481	7%
Somatório	3751		2755		6506	

Fonte: autora desta pesquisa.

A contabilização e a separação de cada constituinte que forma os sintagmas deram-se devido à análise que propomos nesta investigação ser atomística. Tais constituintes foram codificados conforme as suas características morfossintáticas, a partir da presença (“+CN”) ou da ausência (“-CN”) de marca morfológica de número, em concordância com os demais elementos do sintagma, e de acordo com as variáveis divididas nos grupos de fatores apresentados no Quadro 4, abaixo:

Quadro 4 - Variáveis selecionadas para a análise

VARIÁVEL DEPENDENTE:	
Tipo de marca	p (= presença do morfema de plural)
	a (= ausência do morfema de plural)
VARIÁVEIS INDEPENDENTES:	
	Grupo 1: classe gramatical do elemento nominal
	s (= substantivo)
	j (= adjetivo)
	f (= Artigo definido)
	e (= preposição)
	q (= Quantificador)
	t (= Pronome possessivo)
n (= Pronome demonstrativo)	

VARIÁVEIS LINGUÍSTICAS	Grupo 2: posição linear do elemento no sintagma nominal	1 (= posição 1)
		2 (= posição 2)
		3 (= posição 3)
		4 (= posição 4)
	Grupo 3: saliência fônica – processos morfológicos de formação de plural	l (= itens terminados em –l)
		o (= itens terminados em –ão)
		r (= itens terminados em –r)
		c (= itens terminados em –s)
		v (= itens terminados em vogal - plural regular)
		m (= itens terminados em –m)
	Grupo 4: saliência fônica - tonicidade do item lexical	w (=oxítona e monossílabo tônico)
		x (= paroxítona e monossílabo átono)
		y (= proparoxítona)
	Grupo 5: saliência fônica - número de sílabas do item lexical	k (= monossílaba)
		b (= dissílaba)
		d (= mais de duas sílabas)
VARIÁVEIS EXTRALINGUÍSTICAS	Grupo 6: país de origem	G (= Guiné Bissau)
		T (= Timor-Leste)
	Grupo 7: Sexo	M (= feminino)
		H (= masculino)
	Grupo 8: situação no Vestibular da UNILAB-CE	A (= aprovado)
		R (= reprovado)

Fonte: autora desta pesquisa.

Após a codificação dos dados de acordo com as variáveis selecionadas e expressas no Quadro 4, empreendemos a rodada no programa computacional *GoldVarb-X*, a fim de que pudéssemos verificar quais condicionadores exercem maior influência no mecanismo da concordância de número no sintagma nominal, isto é, quais deles favorecem ou não o uso da variante padrão (“+CN”) e não padrão (“-CN”). Com as rodadas, obtivemos o percentual e o peso relativo (doravante PR) das variáveis linguísticas e sociais, escolhidas a partir de Scherre (1988) e estabelecidas para análise. O Quadro 5 apresenta os PR fornecidos pelo programa

GoldVarb-X, os quais indicam o quanto determinadas variáveis foram consideradas relevantes para a análise da variação do fenômeno em questão (uso da concordância nominal de número) e aquela que foi eliminada por não ser relevante.

Quadro 5 - Pesos relativos das variantes dos grupos de fatores, conforme ordem em que apareceram na rodada do programa computacional *goldvarb-x*

Grupos de fatores – Variáveis	Variantes	Código das variantes e PR
Grupo 1: classe gramatical do elemento nominal	Artigo definido	f: 0.555
	Substantivo	s: 0.486
	Adjetivo	j: 0.273
	Preposição	e: 0.588
	Pronome possessivo	t: 0.304
	Pronome demonstrativo	n: 0.451
	Quantificador	q: 0.765
Grupo 2: posição linear do elemento no sintagma nominal	Posição 1	1: 0.473
	Posição 2	2: 0.557
	Posição 3	3: 0.399
	Posição 4	4: 0.169
Grupo 3: saliência fônica - processos morfológicos de formação de plural	Itens terminados em vogal (plural regular)	v: 0.488
	Itens terminados em -r	r: 0.682
	Itens terminados em -m	m: 0.304
	Itens terminados em -s	c: 0.452
	Itens terminados em -l	l: 0.731
	Itens terminados em -ão	o: 0.589
Grupo 4: saliência fônica - tonicidade do item lexical	Paroxítona e monossílabo átono	x: 0.536
	Proparoxítona	y: 0.846
	Oxítona e monossílabo tônico	w: 0.306
Grupo 5: saliência fônica - número de	Monossílaba	k: 0.634
	Dissílaba	b: 0.445

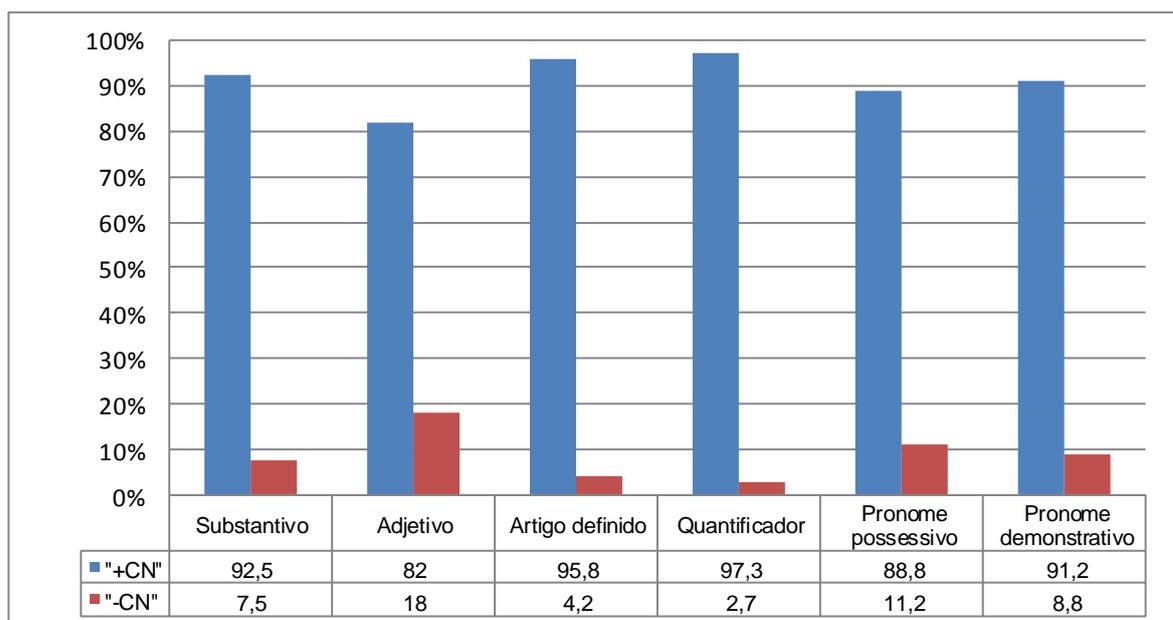
sílabas do item lexical	Mais de duas sílabas	d: 0.426
Grupo 6: país de origem	Guiné Bissau	G:0.533
	Timor-Leste	T: 0.455
Grupo 7: Sexo (eliminado por ser irrelevante)	Homens	H: 0.523
	Mulheres	M: 0.478
Grupo 8: situação no Vestibular da UNILAB-CE	Aprovado	A: 0.632
	Reprovado	R: 0.360

Fonte: autora desta pesquisa.

4.1.1 Variável Classe Gramatical

A primeira rodada dos dados no programa mostrou os quantitativos gerais de presença e ausência de concordância, a partir das variáveis dos grupos de fatores, conforme gráficos a seguir:

Gráfico 2 - Percentual de +CN e -CN, tomando por base a classe gramatical



Fonte: autora desta pesquisa.

Neste estudo, selecionamos 6 classes de palavras cujos usos foram encontrados nos sintagmas nominais das redações analisadas. As demais classes de palavras foram descartadas por não terem aparecido em nossos *corpora*. Acerca da influência da classe gramatical sobre a variação de marcas de plural no sintagma nominal, Poplack (1980a, p. 61, apud SCHERRE, 1988, p. 148), que investigou o espanhol de Porto Rico, concluiu que “os determinantes desfavorecem o cancelamento do -s plural, enquanto os adjetivos e substantivos favorecem-no”. Ainda em relação à classe gramatical, a autora observou que os adjetivos têm menos chances de serem marcados do que os substantivos. Concordamos com a autora no que diz respeito aos adjetivos, uma vez que podemos ver, no Gráfico 2, o percentual de +CN (82%) e -CN (18%) tomando por base as classes de palavras selecionadas. Ademais, é possível observarmos que as classes de substantivos, artigos definidos e quantificadores indefinidos (CASTILHO, 2010) foram as que mais favoreceram o padrão +CN, com percentuais de 92,5%, 95,8% e 97,3%, respectivamente. Isso evidencia que, apesar variedades do português analisadas nesta investigação (PGB e PTL) manterem intenso contato linguístico com as línguas mais faladas em GB e em TL - respectivamente, o crioulo guineense e o tétum, em que não se observam muitas marcas morfológicas de marcação de número plural, sobretudo na classe dos adjetivos -, o PE ainda exerce grande influência no uso dessas variedades.

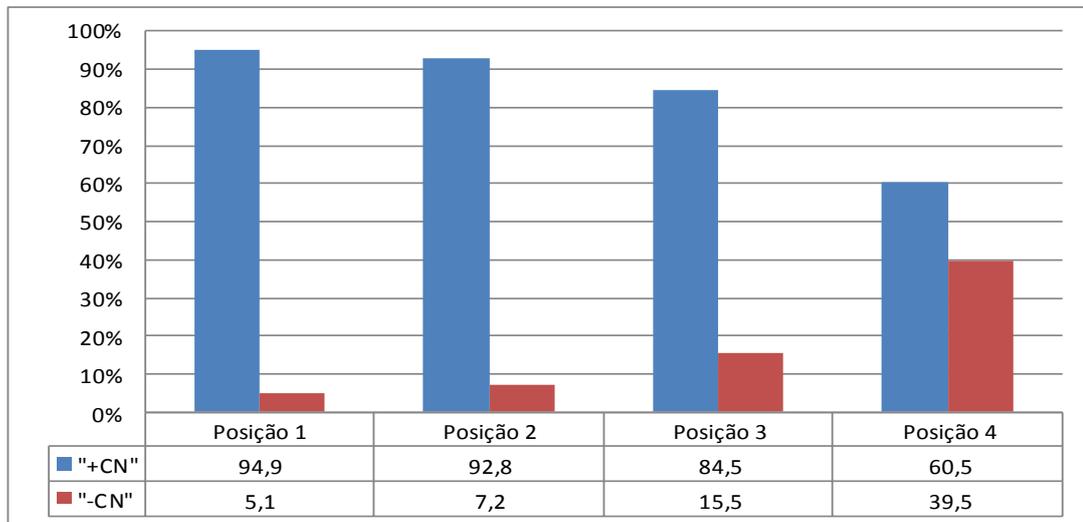
4.1.2 Variável Posição Linear

Em relação à posição linear do constituinte no sintagma nominal, Scherre (1988) aponta:

Todos os trabalhos realizados até o presente momento apresentam, com relação a esta variável, uma conclusão uniforme: a primeira posição do SN é a mais marcada, num índice probabilístico nunca inferior a 0,70; e as demais posições evidenciam um índice baixo de marcas, estabelecendo-se assim uma oposição forte com relação ao que ocorre com o primeiro elemento do SN. No que diz respeito ao comportamento dos elementos da segunda posição em diante, alguns estudiosos encontram uma perfeita linha decrescente. (SCHERRE, 1988, p. 143)

Analisando esta variável, obtivemos os seguintes resultados:

Gráfico 3 - Percentual de +CN e -CN, tomando por base a posição linear do constituinte no sintagma



Fonte: autora desta pesquisa.

A observação do Gráfico 3 nos permite verificarmos que quando o elemento ocupa a primeira posição no sintagma ele tende a receber mais a marca de plural. Esse percentual decresce à medida que o elemento vai se distanciando do início do sintagma, o que corrobora com o que é defendido por Scherre (1988).

A saliente diminuição no uso de marcas de plural que se observa nos constituintes que ocupam a quarta posição do sintagma nominal pode ser ampliada em se tratando de usuários menos escolarizados, os quais, provavelmente, tiveram menos contato com o português em situações formais, sistematizadas e monitoradas de uso, como aquelas proporcionadas pela escola, além do fato do português ser pouco falado fora do espaço escolar e de outros espaços formais, independentemente do tipo de registro e/ou do grau de monitoramento.

Como, neste estudo, tratamos do aspecto “escolarização” pelo viés da situação do candidato no vestibular da UNILAB-CE (aprovado ou reprovado), levantamos a hipótese de que quanto menos “escolarizado” (ou seja, pelo nosso critério, tendo sido reprovado no vestibular), menos o usuário aplicaria a regra de concordância nominal de número em sua redação, visto que o funcionamento do mecanismo de concordância exige mais domínio da norma padrão de uma língua como o português, em que a morfologia é rica e há a existência de muitos elementos flexionáveis. Acerca disso, outra hipótese é a de que os candidatos reprovados

teriam mais influência do contato da L1 do que os aprovados, o que favoreceria, para aqueles, o uso da forma “-CN” e, para estes, o uso da forma “+CN”.

Apresentamos, no Quadro 6, os resultados obtidos no cruzamento dos dados “posição linear” e “situação no vestibular”:

Quadro 6 - Cruzamento dos dados “posição linear” e “situação no vestibular”

	Aprovados		Reprovados	
	Quantitativo	%	Quantitativo	%
Posição 1	+CN = 1395	98%	+CN = 1266	92%
	-CN = 31	2%	-CN = 113	8%
Posição 2	+CN = 1475	97%	+CN = 1233	88%
	-CN = 40	3%	-CN = 169	12%
Posição 3	+CN = 338	87%	+CN = 292	82%
	-CN = 51	13%	-CN = 65	18%
Posição 4	+CN = 14	74%	+CN = 9	47%
	-CN = 5	26%	-CN = 10	53%

Fonte: autora desta pesquisa.

4.1.3 Variável Saliência Fônica

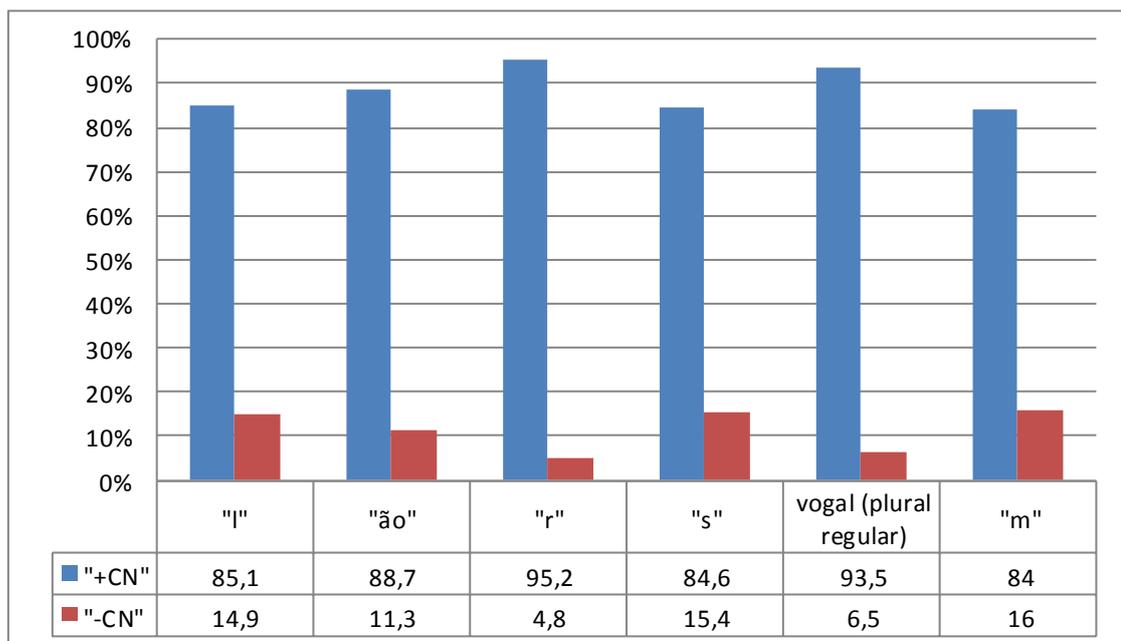
Dando continuidade à descrição dos resultados gerais segundo as variáveis selecionadas, passamos para a variável saliência fônica, a qual, seguindo o que foi feito por Scherre (1988), subdividimos em três dimensões, a saber: 1) processos morfofonológicos de formação de plural, 2) tonicidade do item lexical singular e 3) número de sílabas do item lexical singular.

4.1.3.1 Variável Saliência Fônica – Dimensão Processos

A primeira dimensão, que corresponde aos processos morfofonológicos de formação de plural, diz respeito à diferenciação de material fônico na relação singular/ plural, cuja hipótese, segundo estudos de Scherre (1988), a qual levantamos também nesta pesquisa, é a de que as formas que apresentam maior diferenciação de material fônico na relação singular/plural (os constituintes terminados em *-ão* e em *-l*, por exemplo) tendam a ser mais marcadas do que as que apresentam menor diferenciação (aqueles terminados em vogal, os quais marcam o plural de forma regular, apenas com o acréscimo de *-s*).

A partir da análise global dos dados, confirmamos a hipótese de Scherre (1988) e verificamos um percentual de 93,5% para o padrão “+CN” e de, apenas, 6,5% para “-CN” nos casos em que o plural marcado de forma regular, enquanto, nos casos de maior diferenciação fônica, obtivemos percentuais mais baixos, sendo o menor deles para os elementos terminados em *-m*, com valores de 84% para “+CN” e de 16% para “-CN”, conforme demonstrado no Gráfico 4.

Gráfico 4 - Percentual de +CN e -CN, tomando por base a saliência fônica - dimensão processos



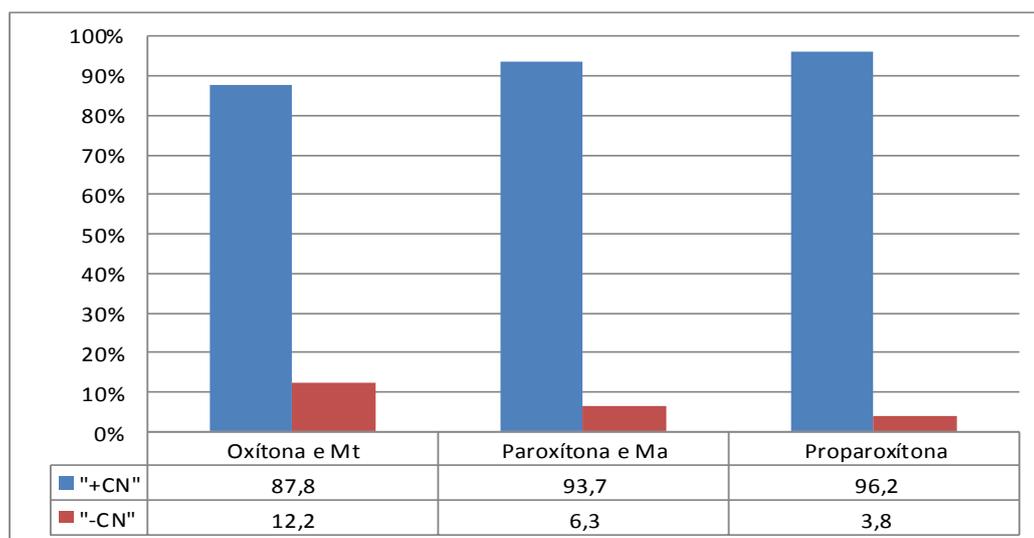
Fonte: autora desta pesquisa.

Outra informação relevante que obtivemos foi a de que os elementos terminados em *-r* totalizaram 95,2% para a forma “+CN” e apenas 4,8% para “-CN”, percentual maior, inclusive, do que aqueles que se referem aos itens terminados em vogal e com plural regular. No entanto, se considerarmos o quantitativo geral de ocorrências e não somente os números percentuais, verificamos que o volume de dados é muito maior para os casos de plural regular (5418 casos, sendo 5065 para +CN e 353 para -CN), ao passo que, para os constituintes terminados em *-r*, o número é significativamente menor (207 casos, sendo 197 para +CN e 10 para -CN), o que torna os 93,5% para “+CN” em itens terminados em vogal e com formação de plural regular ainda um número importante.

4.1.3.2 Variável Saliência Fônica – Dimensão Tonicidade

No que diz respeito à saliência fônica na dimensão tonicidade do item lexical singular, Scherre (1988) levanta a hipótese de que os oxítonos singulares e os monossílabos tônicos, uma vez que sua tonicidade reside na sílaba que vai receber o morfema de plural, favoreçam mais a aplicação da regra do que os paroxítonos e proparoxítonos, que contêm sílabas finais não acentuadas. Nesse aspecto, nossos dados não confirmam a hipótese da autora, conforme apresentado no Gráfico 5:

Gráfico 5 - Percentual de +CN e -CN, tomando por base a saliência fônica – dimensão tonicidade³⁴



Fonte: autora desta pesquisa.

³⁴ Mt = Monossílabo tônico; Ma = Monossílabo átono.

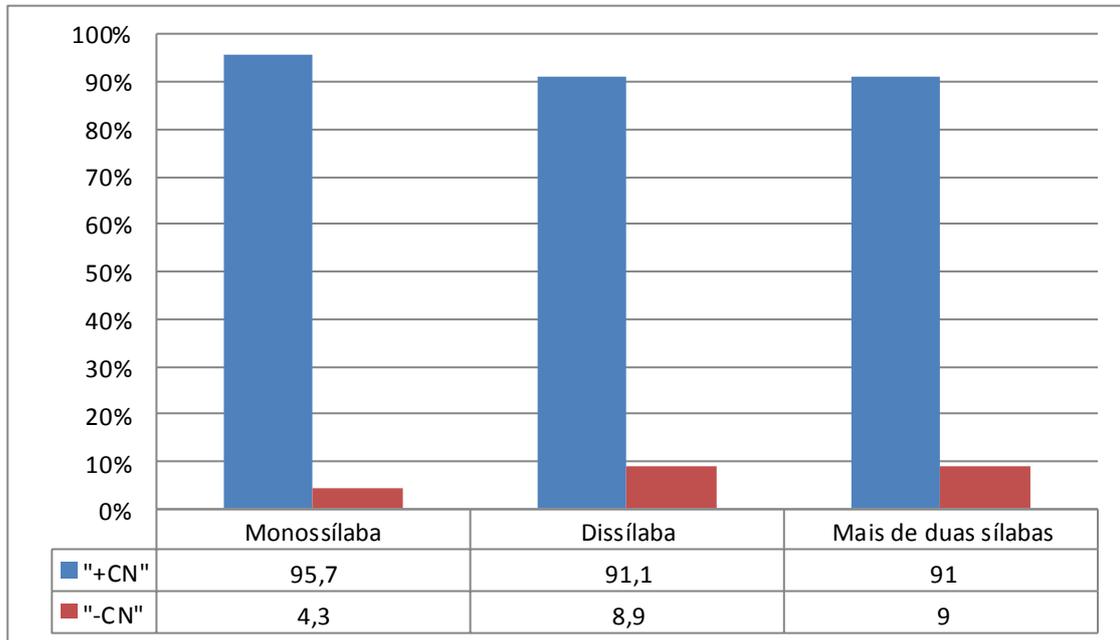
No gráfico, a variável que mais favoreceu a forma “+CN” foi a dos constituintes proparoxítonos. No entanto, à semelhança do que aconteceu com a dimensão Processos - em que os elementos terminados em *-r* totalizaram um percentual maior do que aqueles que se referem aos itens terminados em vogal e com plural regular -, ao verificarmos o quantitativo geral das ocorrências na dimensão Tonicidade, percebemos um volume de dados também muito maior para os casos de palavras proparoxítonas e de monossílabos átonos (4973, sendo 4658 para +CN e 315 para -CN), bem como de oxítonas e monossílabos tônicos (1324, sendo 1163 para +CN e 161 para -CN), do que de palavras proparoxítonas (209, sendo 201 para +CN e 8 para -CN).

4.1.3.3 Variável Saliência Fônica – Dimensão Número de Sílabas

Em relação ao número de sílabas, a rodada com todos os dados apresentou um resultado um pouco divergente daquele encontrado por Scherre (1988). Na pesquisa da autora mencionada, essa variável mostrou-se irrelevante para presença/ausência de concordância entre os elementos do SN, embora tenha obtido percentagens que mostraram uma oposição entre monossílabos (95%), de um lado, e dissílabos (64%) e mais de duas sílabas (58%), de outro (SCHERRE, 1988, p.81).

A rodada que empreendemos com todos os dados de ambos os países em foco considerou, ao contrário do que aconteceu na pesquisa de Scherre, a variável “número de sílabas” como relevante no uso da concordância. Obtivemos, como Scherre (1988), o maior percentual para palavras monossílabas (91%, 2039 casos de um total de 2130), seguido de palavras dissílabas (91,1%, 1612 casos de um total de 1770) e de palavras com mais de duas sílabas (91%, 2371 casos de um total de 2306). Os dois últimos casos, embora não tenham sido tão mais baixos que o primeiro, como foram os de Scherre, acompanham o decréscimo no valor percentual e reiteram que constituintes monossilábicos são mais favorecedores da forma “+CN”, com um peso relativo (doravante PR) de 0.634, enquanto os dissilábicos e os com mais de duas sílabas possuem PR de 0.445 e 0.426, respectivamente. O Gráfico 6, a seguir, apresenta os valores globais obtidos como resultado para essa variável:

Gráfico 6 - Percentual de +CN e -CN, tomando por base a saliência fônica – dimensão número de sílabas



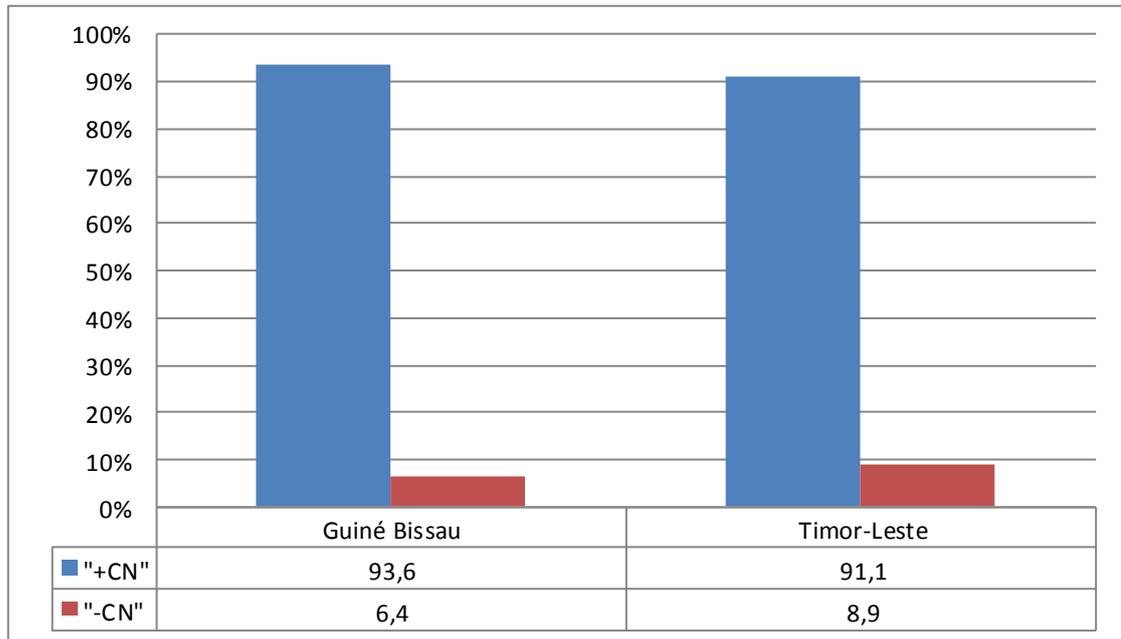
Fonte: autora desta pesquisa.

A partir de agora, aventaremos os resultados globais obtidos a partir das variáveis extralinguísticas.

4.1.4 Variável País de Origem

A primeira variável extralinguística que nos propusemos a analisar foi o país de origem. Esse critério foi utilizado já na separação das redações, na ocasião da coleta de dados realizada na UNILAB-CE. A hipótese que levantamos acerca dessa variável foi a de que os resultados de ambos os países, Guiné Bissau e Timor-Leste, seriam semelhantes. A rodada no programa computacional apresentou uma diferença de pouco mais de dois pontos percentuais para as formas "+CN" (93,6% em GB e 91,1% em TL) e "-CN" (6,4% em GB e 8,9% em TL), o que ressalta a pouca diferença no uso da concordância nos países investigados e coaduna-se com a nossa hipótese.

Apresentamos, no Gráfico 7, os resultados que obtivemos considerando o uso da concordância a partir da variável "país de origem":

Gráfico 7 - Percentual de +CN e -CN, tomando por base o país de origem

Fonte: autora desta pesquisa.

Como dissemos, esse resultado ratifica a hipótese que levantamos inicialmente nesse estudo: que Guiné Bissau e Timor-Leste apresentariam resultados semelhantes quanto ao uso da concordância de número no sintagma nominal. Além disso, os resultados obtidos através da rodada dos dados no programa *GoldVarb-X* ressaltam a pouca diferença que, neste caso, exerce do estatuto da língua no país: L2 em Guiné Bissau e LE em Timor-Leste.

4.1.5 Variável Sexo

A próxima variável extralinguística sobre a qual nos debruçaremos é a variável “sexo”. Nossos resultados mostram, acerca dessa variável, uma diferença muito pequena entre o uso da norma padrão “+CN” por homens (93,2%) e por mulheres (92%), considerando a rodada que fizemos utilizando os dados de Guiné Bissau e de Timor-Leste juntos, como verificamos no Gráfico 8:

Gráfico 8 - Percentual de +CN e -CN, tomando por base o sexo

Fonte: autora desta pesquisa.

Nessa rodada, obtivemos os pesos relativos apresentados pelo *GoldVarb-X*, cuja informação fornecida pelo programa foi a de que o Grupo 7, que corresponde à variável sexo, foi eliminado por não ter sido significativo para aplicação da regra de concordância, com um peso relativo de 0.523, para homens; e 0.478, para mulheres. Acerca do comportamento dessa variável, Scherre (1988), citando Naro e Guy, afirma o seguinte:

Trazendo mais evidências da inoperância da variável Sexo na resolução das questões que estão sob foco, podemos ver que Naro utiliza o comportamento padrão da mulher como uma das evidências de mudança linguística (cf. 1981a, p.86) e Guy utiliza o mesmo raciocínio para concluir sobre variação estável (cf. 1981a, p.198 e 1986, p.11-2), envolvendo o mesmo fenômeno lingüístico: o da concordância de número em Português. (SCHERRE, 1988, p.430)

Tal variável, no estudo da autora, mostrou um grau geral de 0,0 de significância em uma ordem de seleção estatística das variáveis sociais, apresentando relevância, apenas, quando cruzada com outros fatores, como faixa etária e anos de escolarização.

4.1.6 Variável Situação no Vestibular

A última variável extralinguística sobre a qual falaremos isoladamente é a “situação no vestibular”, sobre a qual se faz necessário mencionar alguns aspectos relacionados ao processo seletivo da UNILAB-CE, entre eles os critérios utilizados para aprovação de candidatos vestibular e dos critérios de avaliação das redações produzidas pelos candidatos.

Para alcançarem aprovação no exame³⁵ ao qual estavam se submetendo, os candidatos precisavam apresentar o histórico escolar do Ensino Secundário, o qual geraria a Nota do Ensino Secundário (NES); e escrever a redação de onde retiramos os sintagmas analisados nesta investigação. O cálculo da NES é formado pela média aritmética de cada ano do Ensino Médio (Secundário), a partir da soma de todas as médias finais das disciplinas divididas pelo número total de disciplinas que constam no histórico. Não tivemos acesso ao cálculo da NES de cada aluno do qual selecionamos a redação produzida na ocasião do vestibular, visto que, no momento da coleta dos dados, apenas verificamos o nome dos candidatos - escrito em parte específica da folha de produção de texto - para fazer a separação por sexo (masculino ou feminino) e por situação no vestibular (aprovado ou reprovado), pois eram esses os nossos únicos critérios sociais para a seleção e separação dos textos, além do país de origem (Guiné Bissau ou Timor-Leste). Assim, após a devida separação dos textos pelos critérios sociais, suprimimos o cabeçalho, que continha informações sobre o candidato, como seu nome e seu número de inscrição, antes de digitalizarmos para armazenamento em *pendrive* adquirido para este fim. Feita a digitalização, já não possuíamos a identificação dos escreventes, uma vez que não era esse o nosso propósito, e não acessamos às demais informações que a UNILAB-CE possuía deles.

Em relação à redação propriamente, vimos que ela, em linhas gerais, avalia a expressão escrita do candidato a partir de uma proposta temática. Não tivemos acesso ao tema da redação de 2013, mas, pelo conteúdo das redações, é possível dizermos que, para os candidatos guineenses, foi solicitado que escrevem o porquê

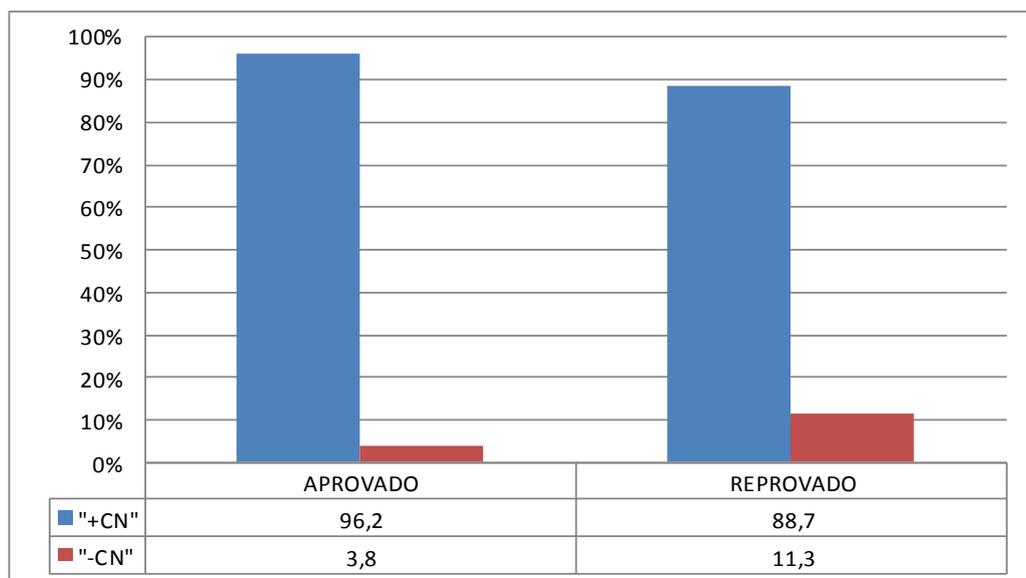
³⁵ As informações referentes ao processo seletivo da UNLAB-CE foram retiradas do edital do Exame de Seleção Vestibular 2013 da referida universidade, disponível em: <http://www.unilab.edu.br/selecao-de-estrangeiros-2013-2-editais/>.

de a língua portuguesa não sair da escrita para a oralidade e, para os candidatos timorenses, que versassem sobre a importância da língua portuguesa em Timor-Leste. Dois temas com teor metalinguístico, que promoviam reflexões acerca da língua e aumentavam ainda mais o grau de monitoramento da situação, o que pode ter se refletido no uso da concordância, visto que, apesar de ter havido ausência de concordância, os números da forma “-CN” foram baixos.

Os critérios utilizados na avaliação das redações envolviam a capacidade de organizar ideias, estabelecer relações, fazer uso de dados ou informações e de elaborar argumentos³⁶. Além disso, avaliava-se domínio de regras gramaticais, como ortografia e pontuação; estratégias de coesão; coerência; e uso da norma culta da língua.

Considerando a composição da NES, salientamos a importância da escolarização para a obtenção, pelo candidato, do êxito no exame vestibular, aspecto que refletiu no resultado do uso da concordância em ambos os países investigados nesta pesquisa, conforme apresentado no Gráfico 9.

Gráfico 9 - Percentual de +CN e -CN, tomando por base a situação no vestibular



Fonte: autora desta pesquisa.

³⁶ Informações extraídas do edital do vestibular 2013 da UNILAB-CE, disponível em: <http://www.unilab.edu.br/wp-content/uploads/2013/07/Edital-68.2013-Processo-Seletivo-de-Estudantes-Estrangeiros-2013.21.pdf>

No Gráfico 9, vemos, por um lado, que os percentuais de uso da forma “+CN” (96,2% para candidatos aprovados e 88,7% para candidatos reprovados) são bastante elevados, com ambos os resultados margeando os 90%. Verificamos, por outro lado, que a forma “-CN” apresenta, entre os aprovados, um percentual menor que 5%, enquanto que entre os reprovados esse número ultrapassa os 10%.

Abordamos, inicialmente, a variável “situação no vestibular” quando analisamos a posição do constituinte no sintagma (Gráfico 3), em que ampliamos a discussão ao relacionarmos o fator “escolarização” (pelo viés da situação do candidato no vestibular da UNILAB-CE, aprovado ou reprovado) ao uso da concordância nominal de número pelo candidato em sua redação. As nossas hipóteses foram as seguintes: a) o funcionamento do mecanismo de concordância exigiria mais domínio da norma padrão de uma língua como o português, em que a morfologia é rica e há a existência de muitos elementos flexionáveis; e b) os candidatos reprovados teriam mais influência do contato da L1 do que os aprovados, o que favoreceria, para aqueles, o uso da forma “-CN” e, para estes, o uso da forma “+CN”.

Conhecendo os critérios utilizados para aprovação de candidatos no vestibular da UNILAB-CE e os critérios de avaliação das redações produzidas por eles na ocasião do exame, assumimos que o resultado apresentado no Gráfico 9 aponta para a confirmação de nossas hipóteses acerca dessa variável.

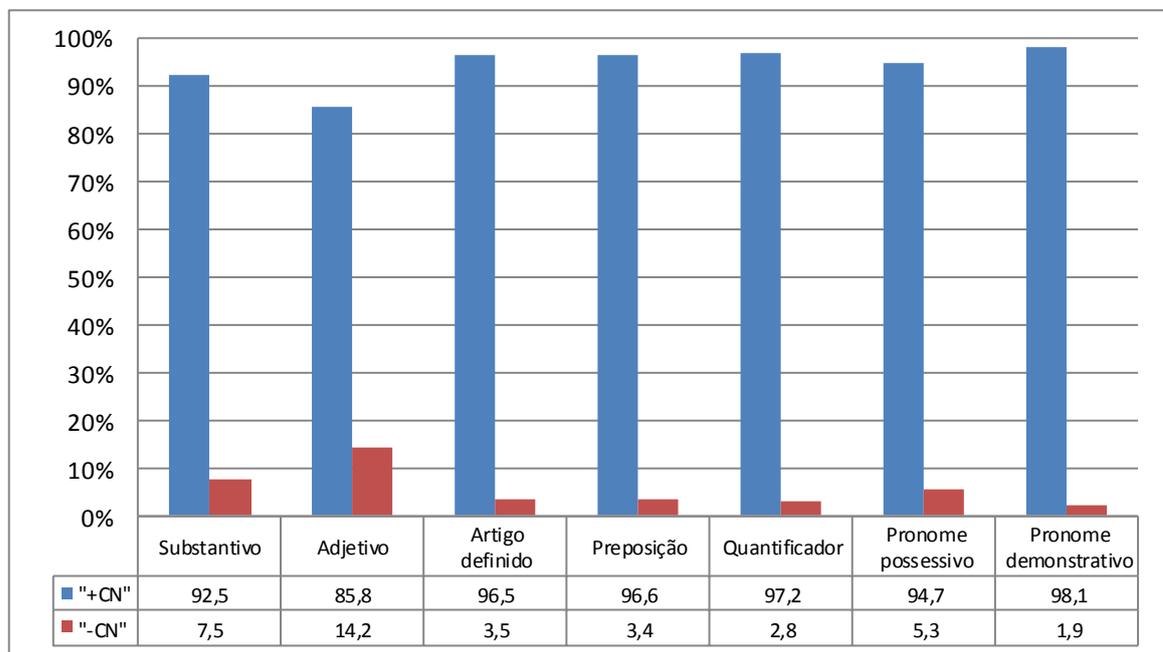
4.2 CONTRASTANDO AS VARIEDADES GUINEENSE E TIMORENSE DO PORTUGUÊS

Neste estudo, buscamos verificar como se dá o uso da concordância nominal de número em redações produzidas por estudantes guineenses e timorenses. Investigações acerca do domínio morfossintático de variedades não europeias do português, o que inclui as variedades africanas e a variedade timorense, foram desenvolvidas por diversos autores (PETTER, 2009; BACELAR DO NASCIMENTO et al., 2008; JON-AND, 2010; ALBUQUERQUE, 2012; BRANDÃO e VIEIRA, 2012a; BRANDÃO e VIEIRA, 2012b; SEDRINS e SILVA, 2017), com alguns dos quais dialogaremos para discutir os resultados apresentados nesta dissertação.

As variáveis linguísticas e extralinguísticas que utilizamos nos permitiram verificar, através da rodada dos dados no programa computacional *GoldVarb-X*, quais grupos de fatores mostraram-se relevantes para o fenômeno investigado. Os dados de Guiné Bissau contabilizaram um total de 3751 (100%) constituintes analisáveis, enquanto os dados de Timor-Leste contabilizaram um total de 2755. Esses dados tiveram os percentuais de uso das formas “+CN” e “-CN” distribuídos com base nas variantes dos grupos de fatores que selecionamos para esta pesquisa. A seguir, apresentamos os gráficos elaborados com base nos resultados obtidos para o uso da concordância nominal de número observado nas redações dos estudantes guineenses e timorenses, contrastando os resultados obtidos nas variedades de cada um dos países em foco.

4.2.1 Variável Classe Gramatical

Gráfico 10 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a classe gramatical



Fonte: autora desta pesquisa.

Como já discutido nos resultados globais, a variável classe gramatical foi considerada significativa para o uso da concordância. Do mesmo modo, das seis

classes de palavras que selecionamos para análise³⁷, os adjetivos foram os que menos receberam a marca morfológica de plural, com percentuais de 85,8% para a forma “+CN” e 14,2% para a forma “-CN” (peso relativo de 0.339). Artigos definidos, preposições, quantificadores e pronomes demonstrativos, por sua vez, foram as classes que mais favoreceram a forma “+CN”, com percentuais de 96,5% (apesar do peso relativo de 0.395), 96,6% (apesar do peso relativo de 0.439), 97,2% (peso relativo de 0.759) e 98,1% (peso relativo de 0.788), respectivamente, o que é exemplificado nos exemplos abaixo, extraídos de nossos *corpora*:

(104) **as** crianças (GHA)

(105) **as** pessoas (GHA)

(106) **os** jovens (GMA)

(107) **dos** guineenses (GMA)

(108) **Das** suas vidas (GMR)

(109) **muitos** jovens (GMR)

(110) **muitos** povos guineenses (GHA)

(111) **várias** razões (GHA)

³⁷ As demais classes de palavras não aparecerem em nossos *corpora*.

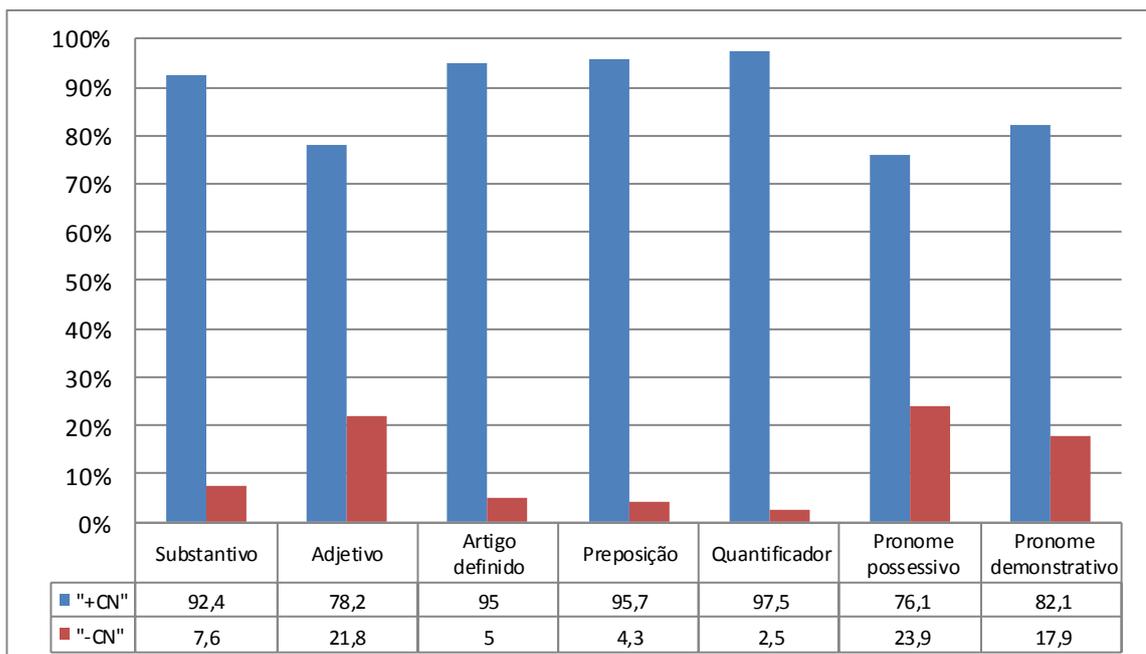
(112) dessas pessoas (GHRA)

(113) essas medidas (GHR)

Sobre a concordância nominal de número no PGB, Sedrins e Silva (2017, p.92) analisaram um total de 912 sintagmas nominais plurais dos quais 865 (95%) aplicaram a concordância, enquanto apenas 47 (5%) não a aplicaram. Apesar dos autores não se referirem especificamente à classe gramatical, o resultado que obtiveram para o uso da forma “+CN” vai ao encontro do nosso, em que a forma “+CN” também foi significativamente mais utilizada.

Os resultados desse grupo de fatores, os quais, com exceção da variante “adjetivo”, superam a marca dos 90% para aplicação da regra de concordância, evidenciam a influência do PE na variedade guineense do português, apesar da já discutida influência do crioulo guineense sobre os usuários.

Gráfico 11 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base a classe gramatical



Fonte: autora desta pesquisa.

Os resultados de Timor-Leste, por sua vez, apresentaram percentuais diferentes daqueles de Guiné Bissau, com a forma “-CN” superando os 20% em algumas classes de palavras: 21,8% nos adjetivos e 23,9% nos pronomes possessivos. No PGB, nenhuma das classes superou a marca dos 15% para a ausência de concordância, tendo sido, inclusive, apenas a classe dos adjetivos a ultrapassar os 10%, enquanto as demais classes ficaram abaixo dessa marca.

Em relação à concordância de número no sintagma nominal do PTL, Albuquerque (2012) defende que “a complexidade do SN é fundamental para se analisar a realização da concordância de número em variedades linguísticas reestruturadas, como é o caso do PTL”. Segundo o autor, caso o sintagma possua menos elementos, como somente o nome ou o determinante com o nome, isso favoreceria a realização da concordância de número. Associando a classe gramatical dos adjetivos, comprovadamente menos marcada morfológicamente com morfema plural (78,2% para “+CN” e 21,8% para “-CN”), com a afirmação de Albuquerque (2012), verificamos que sintagmas formados por determinante, nome e adjetivo corroboram com o que diz o autor em relação ao princípio da simplicidade, conforme demonstrado nos exemplos (114) e (115), extraídos dos *corpora* desta pesquisa:

(114) os professores **preparado** (THR)

(115) dos países **irmão** (TMR)

Além da simplicidade do sintagma, o autor afirma que elementos à direita do núcleo desfavorecem a marcação da concordância, o que também é evidenciado pelos exemplos acima. Assumimos essa característica com o autor, visto que, em geral, os adjetivos aparecem após o nome que constitui o núcleo do sintagma, tendo sua marcação de plural desfavorecida por estas duas razões: estarem à direita do núcleo e comporem sintagmas nominais com estruturas mais complexas, com mais constituintes além do determinante e do nome.

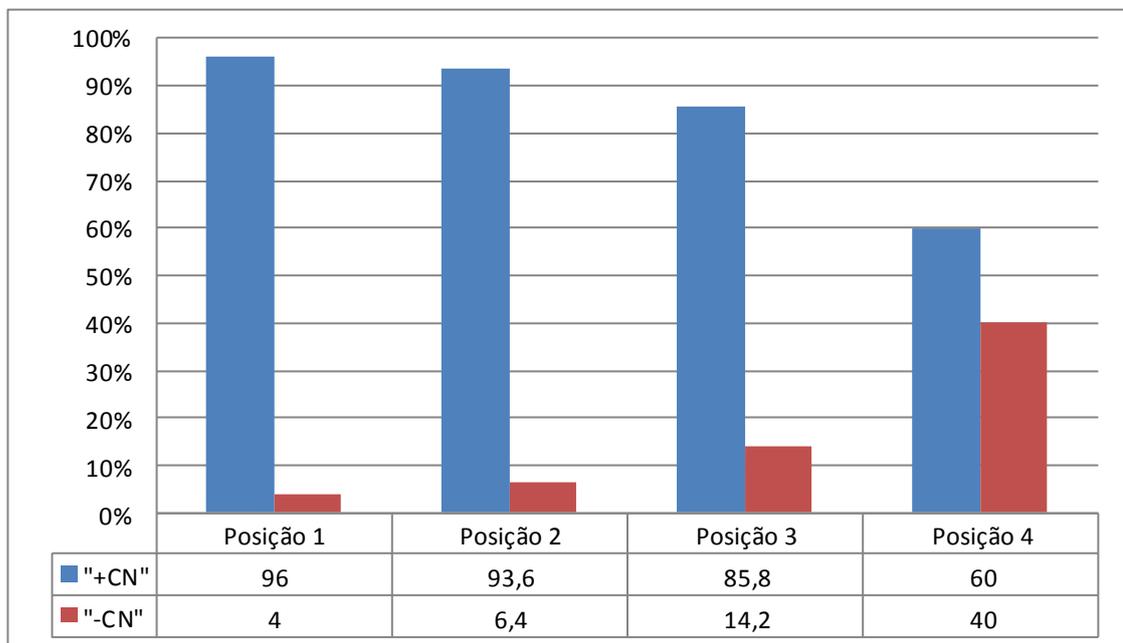
4.2.2 Variável Posição Linear

A posição linear do constituinte no sintagma já se mostrou relevante quando explanamos os resultados globais deste estudo. Analisaremos, agora, como essa variável se comporta em cada um dos países aqui investigados.

Vimos, ao apresentarmos os resultados globais deste estudo, que essa variável mostrou-se bastante significativa, corroborando com o que afirma Scherre (1988), a qual diz que se tratar de uma conclusão uniforme das pesquisas que analisam o fenômeno da concordância nominal de número: a que a primeira posição do SN é a mais marcada e que as demais posições demonstram índices que decrescem significativamente (SCHERRE, 1988, p. 143).

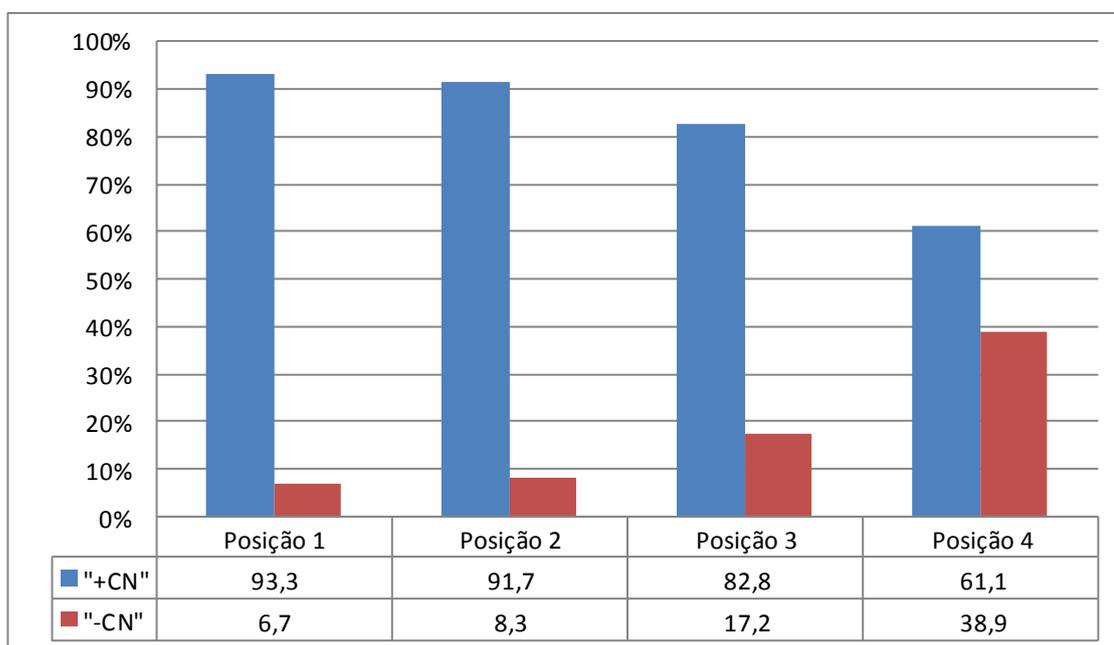
Os Gráficos 12 e 13 contêm os resultados de Guiné Bissau e de Timor-Leste, respectivamente, tomando por base esse grupo de fatores.

Gráfico 12 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a posição linear



Fonte: autora desta pesquisa.

Gráfico 13 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base a posição linear



Fonte: autora desta pesquisa.

Nos gráficos acima, comprovamos a afirmação de Scherre (1988) quanto à posição linear e vemos a ordem decrescente progredindo à medida que o constituinte vai se afastando no início do sintagma, diminuindo em 3%, 8% e 25%, aproximadamente, nos dados de Guiné Bissau, e em 2%, 9% e 21%, aproximadamente, nos dados de Timor-Leste.

A esse respeito, Brandão e Vieira (2012b) observaram, em seu estudo do PB e da variedade africana santomense do português, que a variável posição linear também se mostrou significativa para o favorecimento da marca explícita de plural. Em outro estudo, acerca de três variedades urbanas do português, a saber: o português europeu, o português do Brasil e o português de São Tomé, as autoras afirmaram que

a primeira posição linear, no pré-núcleo, tende a ser mais marcada, o que sugere seja esse o *locus* por excelência da marca. A partir do núcleo em segunda posição, vai decrescendo gradativamente a presença da marca, quer se considerem os menores ou os maiores índices obtidos (BRANDÃO e VIEIRA, 2012a, p. 1045)

Os exemplos de (116) a (120), produzidos por homens guineenses aprovados no vestibular, exemplificam a relevância da variável “posição no sintagma” para o uso da forma “+CN”:

(116) as suas línguas **materna** (GHA)

(117) nos seus próprios **país** (GHA)

(118) nos mercados **guineense** (GHA)

(119) nos países **africano** (GHA)

(120) os países **colonizado** (GHA)

Nos sintagmas constantes em (116) e (117), vemos o constituinte não pluralizado ocupando a quarta posição, enquanto os demais se encontram com marcas morfológicas de plural. Em (118), (119) e (120) a mesma ausência de concordância ocorre com os constituintes que ocupam a terceira posição, enquanto aqueles elementos que ocupam as primeiras e segundas posições estão todos pluralizados.

Sobre o PTL, Albuquerque (2012) coloca a posição linear, a qual ele chama de “ordem”, como um dos fatores estruturais que influenciam a marcação de número, pois, segundo o autor, “os elementos que tendem a ser marcados estão à esquerda do núcleo do SN (linearidade), em sua maioria são determinantes (classe gramatical), e geralmente é o primeiro elemento do SN (ordem)” (ALBUQUERQUE, 2012, p. 5).

Como exemplos da relevância dessa variável, e dos fatores estruturais que favorecem a forma “+CN” em PTL, trazemos os sintagmas abaixo, retirados de nossos *corpora* e produzidos por usuários timorenses aprovados no vestibular:

(121) os **país** membros (THA)

(122) muitos sistemas **português** (THA)

(123) todos os países **lusófono** (THA)

(124) as **língua oficial** (THA)

Os resultados obtidos neste estudo, no que diz respeito à variável posição linear, vão ao encontro do que afirmaram Brandão e Vieira (2012a e 2012b), acerca de uma variedade africana do português (o santomense), o que demonstra semelhança do PGB com essa variedade. Também corroboramos com Albuquerque (2012), em relação ao PTL.

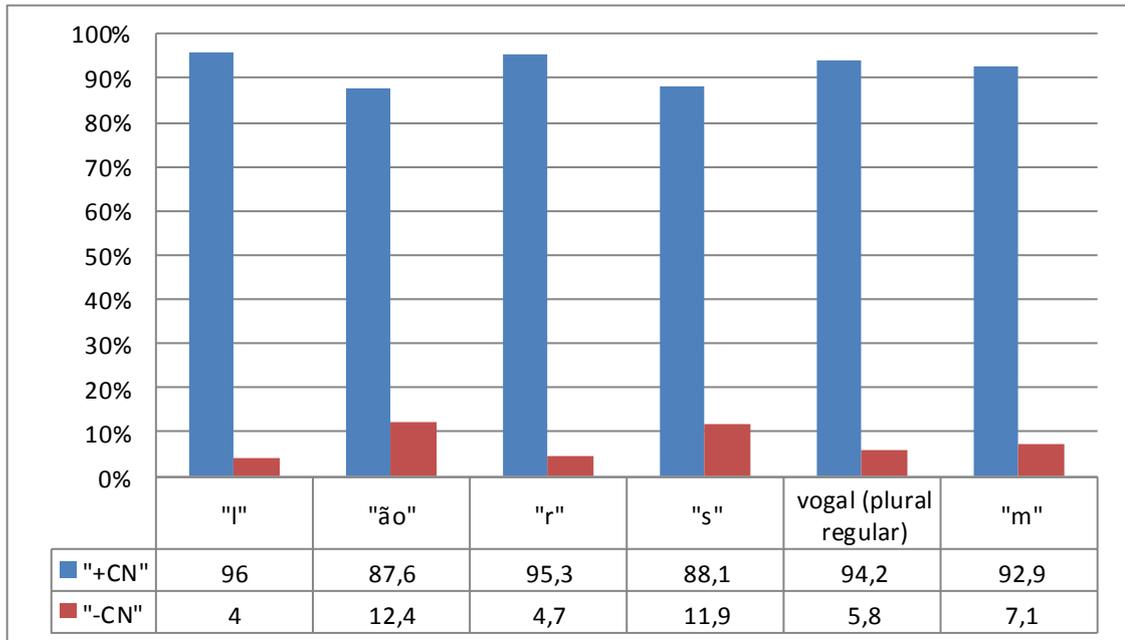
4.2.3 Variável Saliência Fônica – Dimensão Processos

A variável saliência fônica, a qual, neste estudo, dividimos em três dimensões, a exemplo de Scherre (1988), costuma ser inserida em estudos sociolinguísticos. A primeira dimensão que utilizamos para essa variável é composta pelos processos morfofonológicos de formação de plural, dos quais selecionamos sete terminações morfológicas, as quais formam plural de maneiras diversas, para comporem variantes dessa variável. Consideramos, portanto, os itens terminados em -l, os itens terminados em -ão, os itens terminados em -r, aqueles terminados em -s, os terminados em vogal, e, por fim, os terminados em -m.

Como já visto na apresentação dos resultados globais deste estudo, Scherre (1988) levanta a hipótese de que quanto mais diferenciação de material fônico na relação singular-plural, mais haverá marcação de plural. Testamos a hipótese da autora com a rodada dos dados do PGB e do PTL no programa computacional

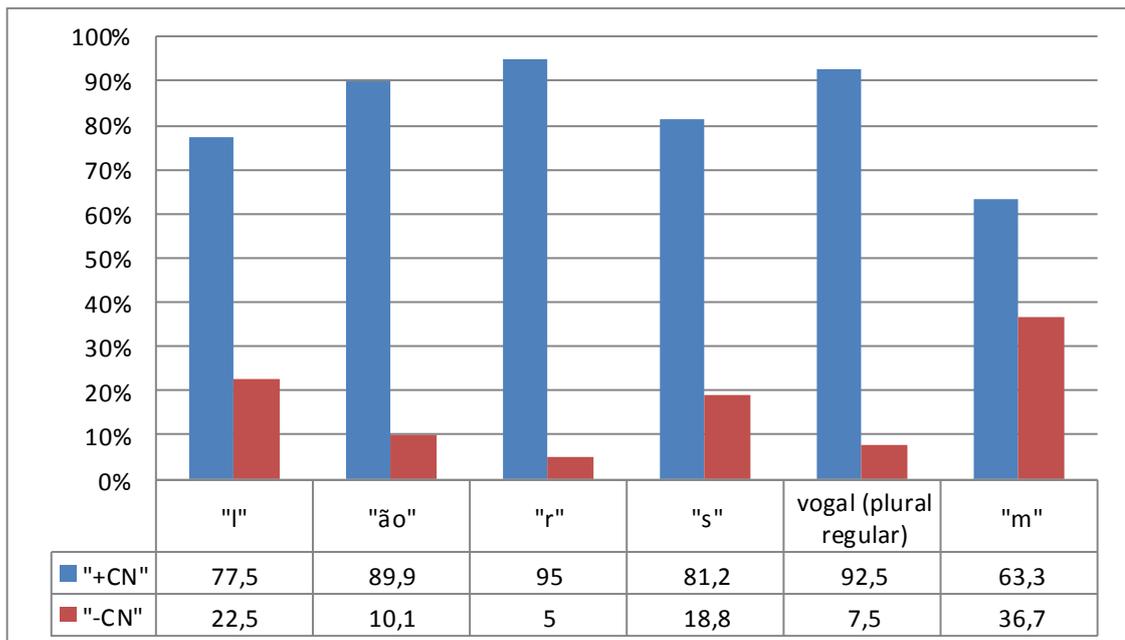
GoldVarb-X, tomando por base essa variável, e apresentamos os resultados nos Gráficos 14 e 15, a seguir.

Gráfico 14 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a saliência fônica – dimensão processos



Fonte: autora desta pesquisa.

Gráfico 15 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base a saliência fônica – dimensão processos



Fonte: autora desta pesquisa.

Como vemos no Gráfico 14, no PGB, os itens terminados em -m, cujo plural se faz com inserção de -s (com ou sem alterações silábicas); aqueles terminados em vogal, que costumam formar plural de maneira regular, apenas com o acréscimo de -s (sem alterações morfofonêmicas); os terminados em -l, com plural formado com a inserção de -s (com ou sem alterações silábicas); e aqueles terminados em -r, cujo plural se faz com inserção de -es (com ou sem alterações silábicas), superaram a marca de 90% para a aplicação da regra de concordância.

No PTL, por sua vez, apenas os itens terminados em vogal e aqueles terminados em -r é que superaram os 90% para a forma “+CN”, com percentuais de 92,5% e 95%, respectivamente, o que contradiz parcialmente a hipótese de Scherre (1988), visto que os itens terminados em vogal, que apresentam menor diferenciação de material fônico na relação singular-plural, tenderiam a ser menos marcados.

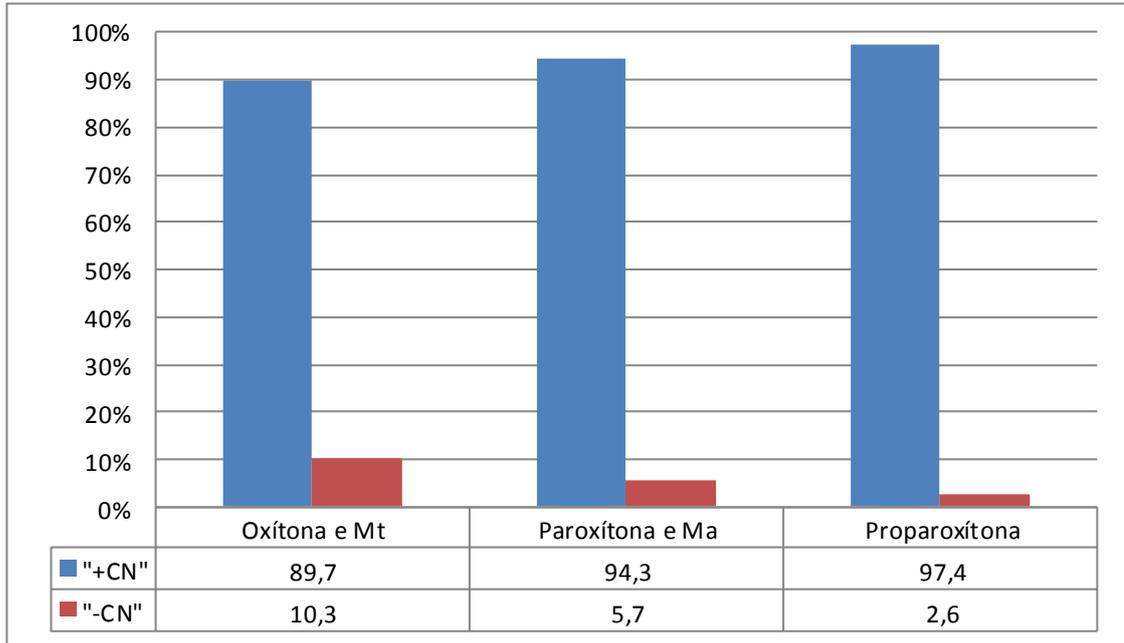
O que se mostra interessante a ser observado é que os itens terminados em -l e em -m comportaram-se de maneira significativamente diferente no PGB e no PTL. Na variedade guineense do português, os constituintes terminados em -l e em -m apresentaram percentuais de 96% e 92,9%, respectivamente, para o uso da forma “+CN”, enquanto na variedade timorense, constituintes com as mesmas terminações apresentaram índices de 77,5% e 63,3%, respectivamente.

4.2.4 Variável Saliência Fônica – Dimensão Tonicidade

A dimensão “tonicidade do item lexical singular”, da variável “saliência fônica”, também foi considerada por Scherre (1988), em sua Reanálise da Concordância Nominal em Português. A autora, como já dissemos na apresentação dos resultados globais para essa dimensão da variável “saliência”, defende que os oxítonos singulares e os monossílabos tônicos, por receberem o morfema de plural na sílaba que contém sua tonicidade, favoreceriam a aplicação da regra de concordância, enquanto os paroxítonos e os proparoxítonos, por não terem sua tonicidade nas sílabas finais, receberiam menos o morfema de plural.

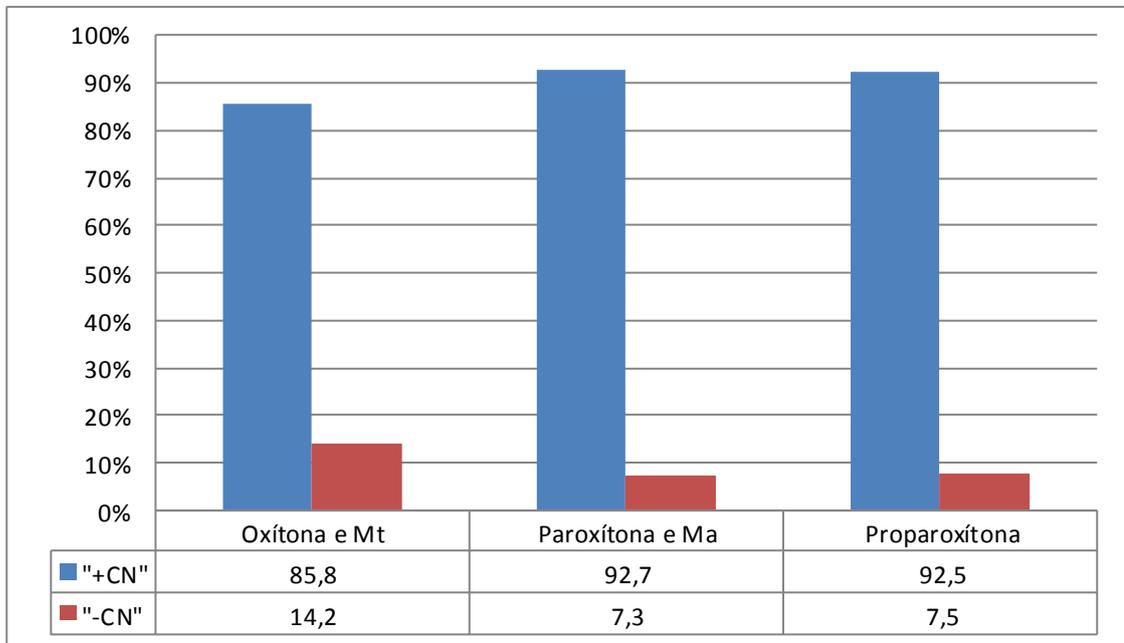
Nossos dados, no entanto, apresentaram resultados que refutam a hipótese de Scherre (1988), conforme vemos nos Gráficos 16 e 17, a seguir.

Gráfico 16 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a saliência fônica – dimensão tonicidade



Fonte: autora desta pesquisa.

Gráfico 17 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a saliência fônica – dimensão tonicidade



Fonte: autora desta pesquisa.

Analisando separadamente os dados do PGB e do PTL, verificamos que os resultados globais mostrados anteriormente já representavam o modo como se

comportam as variedades analisadas neste estudo, no que diz respeito a essa dimensão da variável “saliência fônica”. Tanto na variedade guineense quanto na timorense do português, os constituintes oxítonos e os monossílabos tônicos foram os que menos favoreceram a forma “+CN”, o que vai de encontro ao que é defendido por Scherre (1988). No PGB, constituintes proparoxítonos foram os que mais favoreceram o uso da concordância, e no PTL, os elementos paroxítonos e os monossílabos átonos é que mais contribuíram para esse favorecimento.

Tanto na variedade guineense quanto na variedade timorense do português, o volume de dados de constituintes proparoxítonos foi menor do que os de oxítonos e monossílabos tônicos e menor ainda do que os de elementos paroxítonos. Os números de constituintes para cada variedade estão expressos no Quadro 7, abaixo:

Quadro 7 - Volume de constituintes por país, por número de sílabas e por presença ou ausência de concordância

	Guiné Bissau		Timor-Leste	
	Quantitativo	Total	Quantitativo	Total
Oxítone e monossílabo tônico	+CN = 612	682	+CN = 551	642
	-CN = 70		-CN = 91	
Paroxítone e monossílabo átono	+CN = 2748	2913	+CN = 1910	2060
	-CN = 165		-CN = 150	
Proparoxítone	+CN = 152	156	+CN = 49	53
	-CN = 4		-CN = 4	

Fonte: autora desta pesquisa.

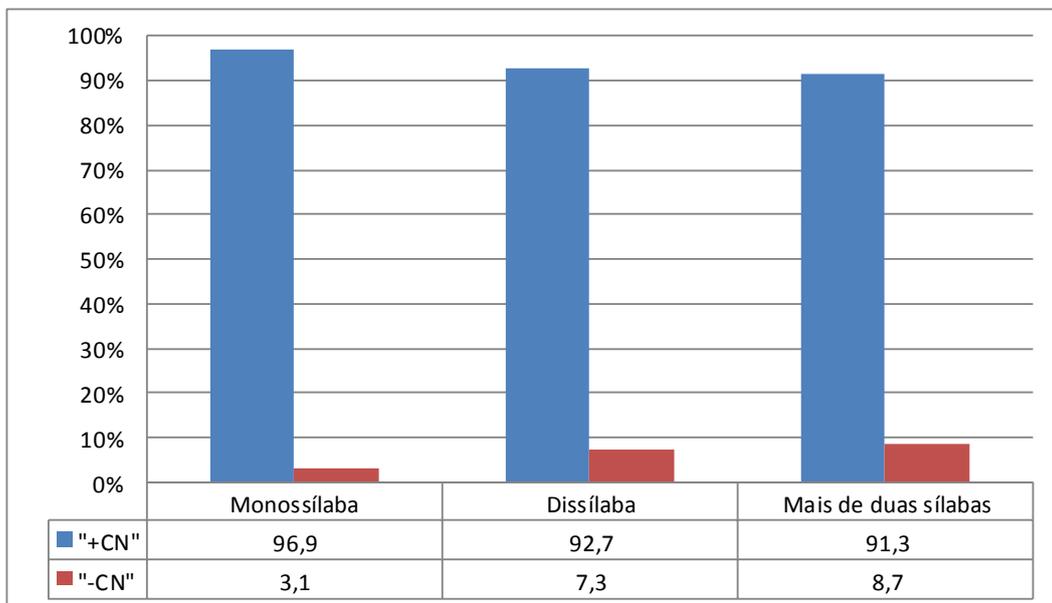
Vale destacar, contudo, que a forma “+CN” foi grande maioria em todas as tonicidades, não sendo o percentual de “-CN” superior a 15% em nenhum dos casos, o que pode indicar o quanto o PE ainda exerce influência no uso da concordância tanto em Guiné Bissau quanto em Timor-Leste.

4.2.5 Variável Saliência Fônica – Dimensão Número de Sílabas

A última dimensão da saliência fônica a ser analisada é a dimensão “número de sílabas”. Essa será, também, a última variável linguística a partir da qual contrastaremos as variedades guineense e timorense do português. No estudo de Scherre (1988) essa dimensão não se mostrou relevante para análise do fenômeno da concordância nominal de número. Optamos por incluí-la em nossa investigação para verificarmos se tal irrelevância se manteria.

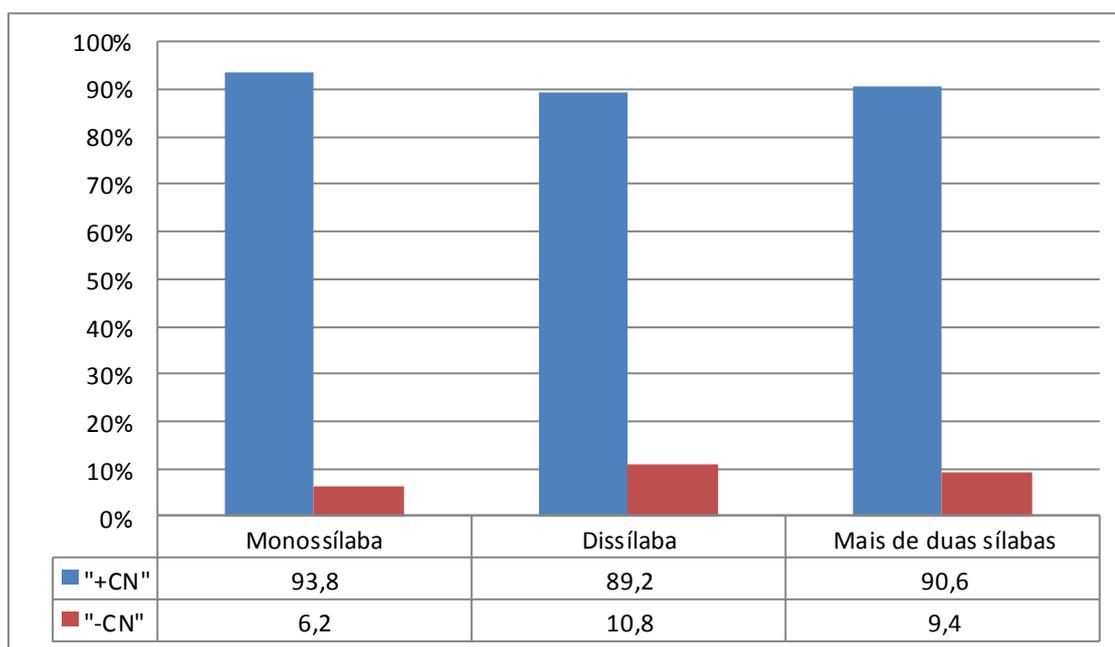
A rodada inicial feita no *GoldVarb-X* com os dados de Guiné Bissau e de Timor-Leste separadamente permitiu que obtivéssemos os quantitativos e percentuais gerais de cada país em relação a essa variável. Os resultados revelados pelo programa computacional estão exibidos nos Gráficos 18 e 19, abaixo, os quais descreveremos e, em seguida, apresentaremos os pesos relativos que nos permitirão compreender por que em Timor-Leste tal grupo de fatores não foi considerado significativo.

Gráfico 18 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a saliência fônica – dimensão número de sílabas



Fonte: autora desta pesquisa.

Gráfico 19 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base a saliência fônica – dimensão número de sílabas



Fonte: autora desta pesquisa.

Como verificamos no Gráfico 18, o percentual de ocorrências da forma “+CN” para elementos monossílabos do PGB demonstra que tal variante favoreceu o uso da concordância, e se distanciou, ainda que pouco, do modo como a regra de concordância foi aplicada nos constituintes paroxítonos e monossílabos átonos e nos proparoxítonos. Já o Gráfico 19, que exhibe os resultados do PTL para essa variável, expõe índices muito próximos para o uso da concordância nas três variantes de grupo de fatores.

A segunda rodada que empreendemos no programa *GoldVarb-X*, ainda com os dados de cada país separadamente, para obtermos os pesos relativos de cada variante dos grupos de fatores, apresentou um resultado diferente para Guiné Bissau e para Timor-Leste, no que diz respeito à dimensão “número de sílabas” da variável saliência fônica. Naquele, por um lado, tal grupo de fatores foi considerado relevante para o uso da concordância, enquanto neste, por outro lado, essa variável foi eliminada por ter sido considerada irrelevante para o uso da concordância. O Quadro 8 apresenta os pesos relativos no PGB e do PTL acerca dessa variável:

Quadro 8 - Pesos relativos da variável saliência fônica na dimensão número de sílabas, por variante e país

Guiné Bissau		Timor-Leste	
Variante	Código da variante e PR	Variante	Código da variante e PR
Monossílaba	k: 0.769	Monossílaba	k: 0.583
Dissílaba	b: 0.363	Dissílaba	b: 0.435
Mais de duas sílabas	d: 0.326	Mais de duas sílabas	d: 0.488

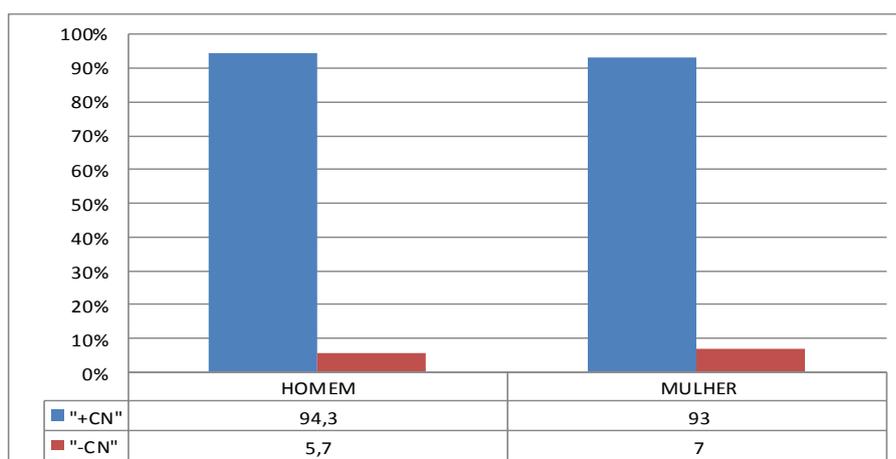
Fonte: autora desta pesquisa.

Tomando por base os pesos relativos apresentados no Quadro 8, concordamos parcialmente com Scherre (1988) quanto à relevância da dimensão “número de sílabas” na análise da variável saliência fônica, visto que em uma das variedades investigadas nesta pesquisa, o PGB, ela foi considerada significativa.

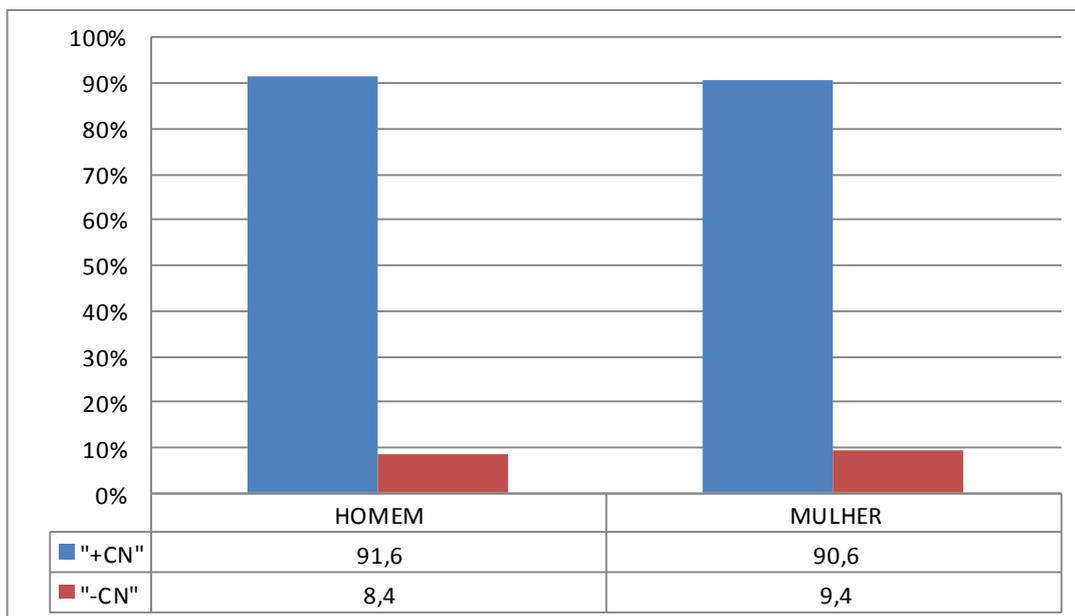
Salientamos, também, que o uso da forma “+CN” foi muito expressivo em ambas as variedades, o que aponta para a grande influência que o PE exerce nas variedades do português em foco.

4.2.6 Variável Sexo

Gráfico 20 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base o sexo



Fonte: Autora desta pesquisa.

Gráfico 21 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base o sexo

Fonte: autora desta pesquisa.

Como vemos, nos Gráficos 20 e 21, há uma diferença muito pequena entre o uso da norma padrão "+CN" por homens e mulheres de ambos os países. Esse dado geral foi confirmado com a rodada em que obtivemos os pesos relativos no *GoldVarb-X*, cuja informação fornecida pelo programa foi a de que o grupo de fatores que corresponde à variável sexo, foi eliminado por não ter sido significativo para aplicação da regra de concordância, com um peso relativo global de 0.523, para homens; e 0.478, para mulheres, conforme apresentamos anteriormente, no Quadro 5. Apresentaremos, agora, no Quadro 9, os pesos relativos mostrados pelo programa computacional na rodada que fizemos com os dados dos países separadamente.

Quadro 9 - Pesos relativos da variável saliência fônica na dimensão número de sílabas, por variante e país

Guiné Bissau		Timor-Leste	
Variante	Código da variante e PR	Variante	Código da variante e PR
Homem	H: 0.535	Homem	H: 0.502
Mulher	M: 0.466	Mulher	M: 0.498

Fonte: autora desta pesquisa.

Acerca do comportamento dessa variável, Scherre (1988), citando Naro e Guy, afirma o seguinte:

Trazendo mais evidências da inoperância da variável Sexo na resolução das questões que estão sob foco, podemos ver que Naro utiliza o comportamento padrão da mulher como uma das evidências de mudança linguística (cf. 1981a, p.86) e Guy utiliza o mesmo raciocínio para concluir sobre variação estável (cf. 1981a, p.198 e 1986, p.11-2), envolvendo o mesmo fenômeno lingüístico: o da concordância de número em Português. (SCHERRE, 1988, p.430)

Tal variável, no estudo da autora, mostrou um grau geral de 0,0 de significância em uma ordem de seleção estatística das variáveis sociais, apresentando relevância, apenas, quando cruzada com outros fatores, como faixa etária e anos de escolarização. Nossos resultados corroboram tanto com a referida autora quanto com os outros estudos verificados por ela.

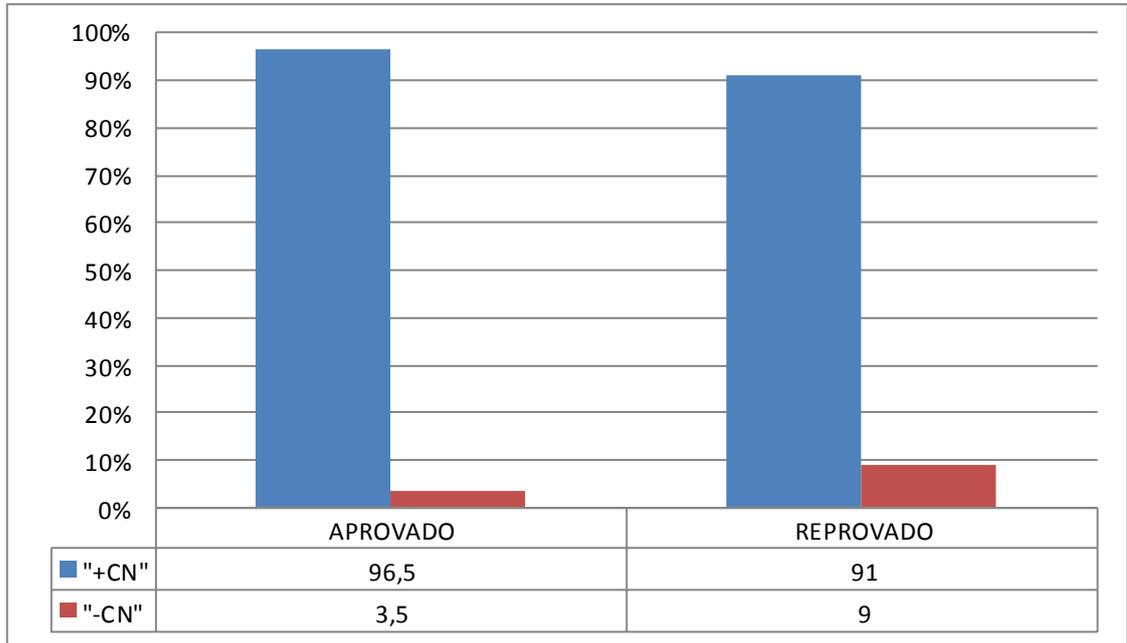
4.2.7 Variável Situação no Vestibular

A situação de “aprovado” ou de “reprovado” em um exame vestibular não costuma fazer parte dos estudos sociolinguísticos. Considera-se, como variável social que se ocupa do fator “educação”, o nível de escolaridade dos usuários. Dito isso, neste estudo equiparamos a variável “situação no vestibular” com a variável “nível de escolaridade”, utilizada comumente em pesquisas de caráter variacionista.

Justificamos essa escolha, como já mencionado anteriormente nesta dissertação, por acreditarmos que tais variáveis aproximam-se. Elencamos duas razões para tal equiparação: 1) o resultado obtido pelos estudantes guineenses e timorenses no vestibular da UNILAB-CE, que nos remete à qualidade na escolarização do candidato; e 2) os critérios para aprovação no vestibular da UNILAB-CE (a composição da NES e a avaliação da redação de onde retiramos os sintagmas que constituem os *corpora* desta investigação), que associam o êxito no exame vestibular à qualidade esperada na escolarização do candidato, aspecto que refletiu no resultado do uso da concordância em ambos os países investigados nesta pesquisa.

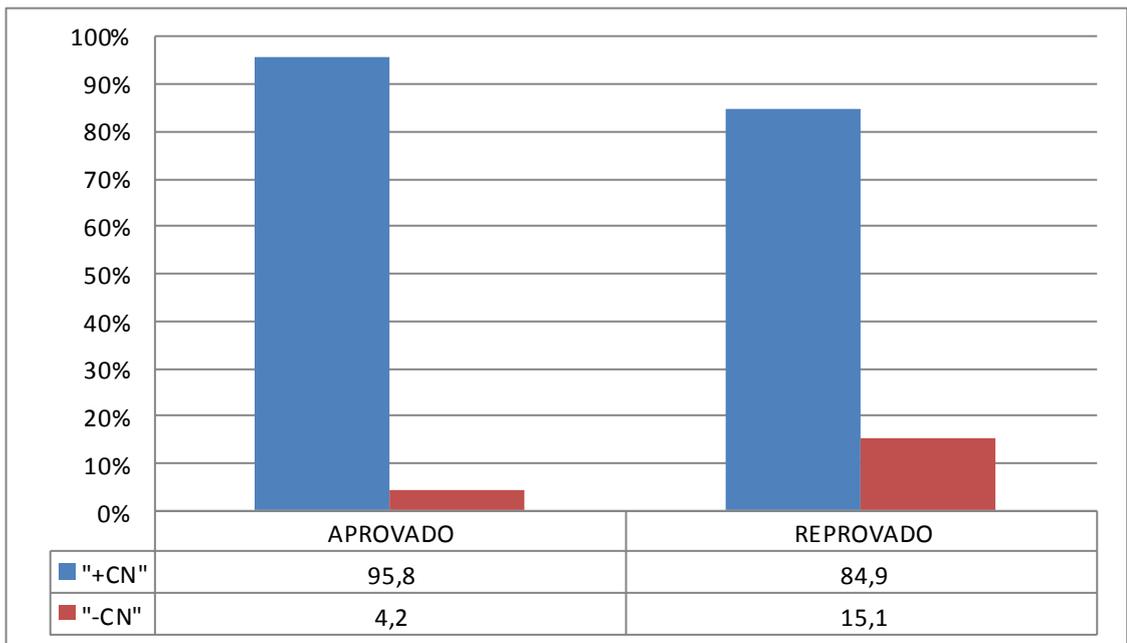
Os Gráficos 22 e 23 apresentam os resultados obtidos com a rodada dos dados de Guiné Bissau e Timor-Leste separadamente para essa variável:

Gráfico 22 - Percentual de +CN e -CN nos dados de Guiné Bissau, tomando por base a situação no vestibular



Fonte: autora desta pesquisa.

Gráfico 23 - Percentual de +CN e -CN nos dados de Timor-Leste, tomando por base a situação no vestibular



Fonte: autora desta pesquisa.

Sintetizamos, no Quadro 10, o cruzamento da variável “situação no vestibular” com a variável “país de origem”:

Quadro 10 - Cruzamento dos dados “país de origem” e “situação no vestibular”, por quantitativo de constituintes analisados e número percentual

	Aprovados		Reprovados	
	Quantitativo	%	Quantitativo	%
Guiné Bissau	+CN = 1728	96,5%	+CN = 1784	91%
	-CN = 62	3,5%	-CN = 177	9%
Timor-Leste	+CN = 1494	95,8%	+CN = 1016	84,9%
	-CN = 65	4,2%	-CN = 180	15,1%

Fonte: autora desta pesquisa.

Como verificamos nos Gráficos 22 e 23 e no Quadro 10, o PGB e o PTL comportam-se de maneira muito semelhante em relação ao uso da concordância, tomando por base, neste caso, a variável situação no vestibular.

No entanto, nos dados de candidatos reprovados do Timor-Leste obtivemos um resultado aproximadamente 6% maior para o favorecimento da marca “-CN”. Essa diferença, ainda que sutil, entre os candidatos do Timor-Leste reprovados no vestibular da UNILAB-CE, pode apontar para o fato de que o estatuto de língua estrangeira do português no Timor-Leste talvez tenha permitido que o tétum exercesse mais influência no uso do PTL do que o crioulo guineense exerceu no uso do PGB, considerado, por sua vez, como segunda língua na Guiné Bissau.

O estatuto de segunda língua ou de língua estrangeira está associado à relação que os usuários estabelecem com a língua no país, bem como com o próprio espaço que a língua terá nos ambientes formais de uso. Nesse cenário, a reprovação do candidato no vestibular da UNILAB-CE, situação que podemos associar ao uso da concordância, sugere que a qualidade na escolarização não foi muito eficiente e/ou que, se não for através da educação formal e de situações

sistematizadas que exijam dos candidatos o uso do português, bem como o contato com a norma padrão e o domínio do uso da concordância, haverá uma diminuição do uso da forma “+CN”.

Retomamos, aqui, os estudos de Gonçalves (2012, apud SEDRINS e SILVA, 2017) e os de Brandão e Vieira (2012ab), os quais colocam a variável nível de escolaridade como sendo extremamente relevante para a ocorrência e a produtividade dos padrões de variação no âmbito da concordância nominal de número.

4.3 QUADROS-RESUMO DOS PESOS RELATIVOS POR GRUPO DE FATORES E POR VARIANTE PARA O USO DO PGB E DO PTL

Quadro 11 - Pesos relativos das variantes dos grupos de fatores – país Guiné Bissau

Grupos de fatores – Variáveis	Variantes	Código das variantes e PR
Grupo 1: classe gramatical do elemento nominal	Artigo definido	f: 0.395
	Substantivo	s: 0.556
	Adjetivo	j: 0.339
	Preposição	e: 0.439
	Pronome possessivo	t: 0.512
	Pronome demonstrativo	n: 0.788
	Quantificador	q: 0.759
Grupo 2: posição linear do elemento no sintagma nominal	Posição 1	1: 0.488
	Posição 2	2: 0.558
	Posição 3	3: 0.339
	Posição 4	4: 0.101
Grupo 3: saliência fônica - processos morfológicos de formação de plural	Itens terminados em vogal (plural regular)	v: 0.478
	Itens terminados em -r	r: 0.736
	Itens terminados em -m	m: 0.460
	Itens terminados em -s	c: 0.518

	Itens terminados em -l	l: 0.932
	Itens terminados em -ão	o: 0.532
Grupo 4: saliência fônica - tonicidade do item lexical	Paroxítona e monossílabo átono	x: 0.523
	Proparoxítona	y: 0.917
	Oxítona e monossílabo tônico	w: 0.280
Grupo 5: saliência fônica - número de sílabas do item lexical	Monossílaba	k: 0.769
	Dissílaba	b: 0.363
	Mais de duas sílabas	d: 0.326
Grupo 6: Sexo (eliminado por ser irrelevante)	Homens	H: 0.535
	Mulheres	M: 0.466
Grupo 7: situação no Vestibular da UNILAB-CE	Aprovado	A: 0.622
	Reprovado	R: 0.388

Fonte: autora desta pesquisa.

Quadro 12 - Pesos relativos das variantes dos grupos de fatores - país Timor-Leste

Grupos de fatores – Variáveis	Variantes	Código das variantes e PR
Grupo 1: classe gramatical do elemento nominal	Artigo definido	f: 0.637
	Substantivo	s: 0.447
	Adjetivo	j: 0.254
	Preposição	e: 0.644
	Pronome possessivo	t: 0.168
	Pronome demonstrativo	n: 0.303
	Quantificador	q: 0.790
Grupo 2: posição linear do elemento no sintagma nominal	Posição 1	1: 0.463
	Posição 2	2: 0.556
	Posição 3	3: 0.442

	Posição 4	4: 0.207
Grupo 3: saliência fônica - processos morfológicos de formação de plural	Itens terminados em vogal (plural regular)	v: 0.504
	Itens terminados em -r	r: 0.581
	Itens terminados em -m	m: 0.212
	Itens terminados em -s	c: 0.334
	Itens terminados em -l	l: 0.561
	Itens terminados em -ão	o: 0.594
Grupo 4: saliência fônica - tonicidade do item lexical	Paroxítona e monossílabo átono	x: 0.537
	Proparoxítona	y: 0.661
	Oxítona e monossílabo tônico	w: 0.372
Grupo 5: saliência fônica - número de sílabas do item lexical (eliminado por ser irrelevante)	Monossílabo	k: 0.583
	Dissílabo	b: 0.435
	Mais de duas sílabas	d: 0.488
Grupo 6: Sexo (eliminado por ser irrelevante)	Homens	H: 0.502
	Mulheres	M: 0.498
Grupo 7: situação no Vestibular da UNILAB- CE	Aprovado	A: 0.633
	Reprovado	R: 0.330

Fonte: autora desta pesquisa.

Com a utilização do programa computacional *Goldvarb-X*, é possível que obtenhamos informações diversas acerca do fenômeno investigado. Dentre essas informações, os pesos relativos representam a verdadeira relevância de determinada variável, pois, paralelas às rodadas que fazemos no *software*, ele realiza outros cruzamentos que indicam, inclusive, se devemos considerar as informações acerca de determinada variável ou se devemos eliminá-la por não ser

significativa. Nesses casos, ao apresentar alguns dos resultados, o próprio *Goldvarb-X* informa que determinada variável foi eliminada.

Os Quadros 11 e 12, os quais denominamos quadros-resumo, que elaboramos na presente seção, apresentam os pesos relativos obtidos para cada uma das variantes por grupo de fator. Neles, destacamos aquelas que podemos considerar significativas. Das variáveis eliminadas por serem irrelevantes, o PR não foi destacado.

5 CONSIDERAÇÕES FINAIS

Como vimos no início desta dissertação, o português do Brasil é uma variedade do português já bastante consolidada e difundida. Vimos, em adição, que outras variedades não europeias do português também precisam ser estudadas a fim de que se conheça melhor sua realidade linguística. Mencionamos a CPLP que foi criada com o propósito de integrar e estabelecer parcerias entre os países que têm o português como língua oficial, a saber: Angola, Brasil, Cabo Verde, Guiné-Bissau, Moçambique, Portugal, São Tomé e Príncipe, e, ainda que tardiamente, o Timor-Leste. Salientamos, também, que, apesar dessa proposta de integração entre os países lusófonos, ainda há, atualmente, grande lacuna e latente desequilíbrio no prestígio de algumas variedades do português, sobretudo as africanas e a asiática, do Timor-Leste.

Um exemplo desse problema são os diversos estudos no âmbito da concordância nominal, por exemplo, que têm sido realizados ao longo dos anos no PB e no PE, o que não se equipara aos estudos relativos às outras variedades. Nossa pesquisa, nesse cenário, buscou salientar a relevância de estudos que tratem de aspectos morfossintáticos de variedades do português, a fim de compreender os processos que provocam uma “polarização sociolinguística”, conforme dito por Brandão (2016), a qual contribui para a formação identitária tanto de países lusófonos da África quanto do país asiático de Timor-Leste.

Dentre os estudos nos quais nos embasamos para empreendermos a investigação que aqui se propôs, destaca-se o estudo precursor desenvolvido por Scherre (1988), em que a autora realizou uma reanálise da concordância nominal em português, com o qual os resultados acerca das variedades do português analisadas nesta pesquisa, o PGB e o PTL, tomando por base as variáveis que selecionamos, foram comparados parcialmente. A comparação parcial deu-se pelo fato de não termos utilizado, aqui, até por uma questão de escassez de tempo, todas as variáveis aplicadas pela autora. Todavia, os grupos de fatores que utilizamos nos permitiram traçar comparações entre os resultados da referida autora e os nossos.

Esclarecemos, oportunamente, que, conforme explicitado ainda nas palavras iniciais desta dissertação, tínhamos como objetivo, no início deste estudo, analisar tanto a variação da concordância nominal de número quanto a de gênero no PGB e

no PTL, no entanto, como também já informado, devido à escassez de tempo, optamos por fazer a análise somente na perspectiva da concordância nominal de número, deixando a análise do uso da concordância nominal de gênero para estudos futuros, de modo que pudéssemos nos debruçar melhor sobre este fenômeno e realizar uma análise pormenorizada dessa questão.

Retomando os objetivos deste estudo, em que nos propusemos a investigar o uso da concordância nominal de número em redações produzidas em português por estudantes guineenses e timorenses no âmbito do exame vestibular da UNILAB-CE, verificamos que o intenso contato linguístico do PGB e do PTL com o crioulo guineense e com o tétum, respectivamente, influencia para que o uso da concordância seja uma regra variável, de acordo com os critérios estabelecidos por Labov (2003). O resultado global apresentado, em que vimos um percentual de 93,6% para a forma “+CN”, em Guiné Bissau, e de 91,1% em Timor-Leste, confirma a hipótese geral deste trabalho, em que foi dito que: 1) devido ao fato da situação de uso da língua ser um contexto monitorado (a situação de um exame), não se esperavam muitas ausências de concordância nos dados analisados; e 2) a forte influência do PE, apesar do intenso contato do PGB com o crioulo guineense e do PTL com o tétum - línguas consideradas, neste estudo, como maternas dos falantes guineenses e timorenses, respectivamente, em que não se observa, na maioria dos usos, marcas explícitas de número em alguns dos elementos no interior do sintagma nominal (CASTRO, 2013; ALBUQUERQUE, 2012) -, que permitiu que houvesse, tanto no PGB quanto no PTL, que a ausência de concordância não fosse tão frequente nos dados escritos analisados.

Um de nossos objetivos específicos foi, também, compreender a possível interferência da morfossintaxe do tétum e do guineense no PTL e no PGB, respectivamente, em particular, no âmbito da concordância nominal de número, tendo em mente a aquisição do português como L2 e sua coexistência com as línguas maternas (L1) desses países. Acerca desse objetivo, vimos que o estatuto de língua é diferente para ambos os países em foco, sendo considerada L2 em Guiné Bissau e LE em Timor-Leste. Essa diferença, contudo, não foi determinante para que os resultados dos países diferissem significativamente, e os resultados obtidos evidenciam índices percentuais, como já mencionado nessas considerações finais, de 93,6% para a forma “+CN”, em Guiné Bissau, e de 91,1%, em Timor-Leste, o que caracteriza um fenômeno variável. Os resultados foram obtidos através de

rodadas dos dados no programa computacional *GoldVarb-X*, o qual forneceu dados numéricos que permitiram uma análise estatístico-comparativa do fenômeno em questão, nas variedades do português investigadas neste estudo.

Nos dedicamos a contrastar, também, a morfologia flexional de número no crioulo guineense e no tétum com a do PGB e do PTL encontrada nos dados em análise, tendo em mente possíveis interferências da L1 na L2. Quanto a isso, havíamos levantado a hipótese de que o contato linguístico em Guiné-Bissau e Timor Leste, a aquisição do português como L2 e sua coexistência com as línguas maternas (L1) desses países, respectivamente, o crioulo guineense e o tétum, fariam com que a aplicação da regra de concordância fosse um fenômeno variável. Verificamos, acerca disso, que, embora não se observem naquelas uma diversidade de marcas explícitas de morfologia flexional de número, que exijam maior domínio do mecanismo de concordância; nestas, por sua vez, ainda que a morfologia flexional seja rica e complexa, o uso da concordância seguiu os padrões da norma culta, sugerindo forte influência do PE. Confirmamos nossa hipótese, seguindo os limites percentuais estabelecidos por Labov (2003), de 100% para regras categóricas; de 95 a 99% para semicategóricas; e de 5 a 95% para regras variáveis, apesar dos índices superando os 90% terem sido bastante elevados para o uso padrão da regra de concordância.

Dando continuidade aos objetivos que traçamos para esta investigação, buscamos verificar fatores de natureza linguística e extralinguística que poderiam favorecer o uso da concordância nominal de número no PGB e no PTL. Nossa hipótese, quanto a isso, era de que de, conforme defendido por Labov ([1972], 2008), haveria padrões observáveis na variação, ou seja, ela não ocorreria de forma aleatória, e os fatores de natureza linguística e extralinguística favoreceriam mais ou menos o uso da concordância nominal de número no PGB e no PTL. Nossos resultados para esse objetivo e acerca dessa hipótese consideraram relevantes para a ocorrência da variação a maioria das variáveis que selecionamos para este estudo, com exceção, nos resultados globais e no resultado individual de Guiné Bissau, da variável “sexo”; e nos resultados individuais de Timor-Leste, a variável “sexo” e a variável “saliência fônica” em sua dimensão número de sílabas.

No primeiro caso, dos resultados globais, em que empreendemos a rodada dos dados dos dois países simultaneamente no programa computacional *GoldVarb-X*, a variável sexo foi eliminada por não ter sido significativa para aplicação da regra

de concordância, com um peso relativo de 0.523, para homens; e 0.478, para mulheres. Na rodada individual dos dados de Guiné Bissau, obtivemos um resultado semelhante, em que essa variável também foi eliminada por não ter sido relevante para a ocorrência do fenômeno em questão, com peso relativo de 0.535 para homens e de 0.466 para mulheres. Ao empreendermos a rodada individual dos dados de Timor-Leste, obtivemos um peso relativo de 0.502 para homens e 0.498 para mulheres, o que fez com que essa variável fosse eliminada, da forma como ocorreu com Guiné Bissau, por não ter sido significativa para a frequência de produtividade do fenômeno analisado.

Ainda em relação ao Timor-Leste, a rodada individual dos seus dados no *GoldVarb-X* apresentou outro resultado interessante: a variável linguística saliência fônica, em sua dimensão “número de sílabas”, também foi eliminada, a exemplo do que aconteceu com a variável extralinguística “sexo”, por não ter se mostrado relevante para o uso da concordância nominal de número. O peso relativo das variantes dessa variável foi, para constituintes monossílabos, 0.583; para constituintes dissílabos, 0.435; e para constituintes com mais de duas sílabas, 0.488. As mesmas variantes desse grupo de fatores, aplicadas aos dados de Guiné Bissau, apresentaram pesos relativos de 0.769, 0.363 e 0.326, respectivamente, tendo sido consideradas significativas para a ocorrência da variação no uso da concordância.

As demais variáveis linguísticas e extralinguísticas foram consideradas relevantes para a ocorrência e a produtividade do fenômeno em questão, e seus respectivos pesos relativos estão apresentados nos quadros-resumo constantes no último capítulo desta dissertação.

Realizamos, também, algumas rodadas efetuando cruzamentos entre variáveis, e uma delas nos forneceu resultados bastante expressivos: aquela em que cruzamos os grupos de fatores “posição linear” (variável linguística) e “situação no vestibular” (variável extralinguística). Lembramos que equiparamos, nesta pesquisa, a situação no vestibular ao fator “escolaridade”, e verificamos que, entre os candidatos aprovados, o uso da forma “+CN” foi maior, bem como a concordância aplicada a constituintes mais distantes do início do sintagma. Isso evidencia que quanto mais escolarizado é o falante, considerando que a escolarização foi mais efetiva nos candidatos aprovados do que nos reprovados, maiores serão os índices de aplicação da regra de concordância e maior deve ser o domínio que ele possui dos mecanismos de concordância nominal de número.

Após a análise global e contrastiva dos resultados, alcançamos o último objetivo específico traçado para esta investigação: o de analisar os padrões de concordância nominal de número verificados no PGB e no PLT, levando em conta os tipos de regras linguísticas propostas por Labov (2003) à luz da Sociolinguística Variacionista. Nossa hipótese era a de que evidências encontradas na análise sociolinguística poderiam comprovar que o tipo de regra que se aplica aos padrões de concordância de número no PGB e no PTL possuem característica de uma regra variável, hipótese essa que foi confirmada através dos resultados que obtivemos.

Por fim, tendo concluído a pesquisa que nos empenhamos em desenvolver, deixamos, também, como sugestão para estudos futuros que possam ser desenvolvidos, o fenômeno da concordância nominal de gênero nas variedades guineense e timorense do português, bem como o estudo de outros fenômenos analisáveis nas variedades desse “grande corpo espalhado pelo mundo” que é a língua portuguesa.

REFERÊNCIAS

ALBUQUERQUE, D. B. de. **Esboço morfossintático do português falado em Timor-Leste**. Brasília: Moderna språk. Universidade de Brasília. 2012.

_____. **Esboço gramatical do tétum prasa**: língua oficial de Timor-Leste. Dissertação de Mestrado. Brasília-DF: Universidade de Brasília. Instituto de Letras. Departamento de Linguística, Português e Línguas Clássicas. Programa de Pós-Graduação em Linguística. 2011. 194p.

ALMEIDA, N. C. H. **LÍNGUA PORTUGUESA EM TIMOR-LESTE: ENSINO E CIDADANIA**. Dissertação de Mestrado em Língua e Cultura Portuguesa. UNIVERSIDADE DE LISBOA. FACULDADE DE LETRAS. DEPARTAMENTO DE LÍNGUA E CULTURA PORTUGUESA. 2008.

BACELAR DO NASCIMENTO et al. **Corpus África**: as cinco variedades africanas do português. XXIII Encontro Nacional da Associação Portuguesa de Linguística, Lisboa, APL, p. 373-384, 2008.

BRANDÃO, S. F. Variação e o estatuto de variedades do português. **Diadorim**, Rio de Janeiro, Especial 2016, p. 83-104. 2016.

BRANDÃO, S. F. Concordância nominal em três variedades do português: resultados gerais, novas indagações. **Cuadernos de la alfa**, nº 7. ISSN 2218-0761. 2015

BRANDÃO, S. F.; VIEIRA, S. R. Concordância nominal e verbal: contribuições para o debate sobre o estatuto da variação em três variedades urbanas do português. **Alfa**, São Paulo, v. 53, n. 3, p. 1035-1064, 2012a.

BRANDÃO, S. F.; VIEIRA, S. R. A concordância nominal e verbal no Português do Brasil e no Português de São Tomé: uma abordagem Sociolinguística. **PAPIA** ISSN 0103-9415 e ISSN 2316-2767. p. 7-39, 2012b.

CASTILHO, A. T. de. **Nova Gramática do Português Brasileiro**. São Paulo: Contexto, 2010.

CASTRO, P. P. **As construções interrogativas, de tópico e de foco na língua crioula de Guiné-Bissau**. 2013. Dissertação (Mestrado em Linguística) - Universidade Federal do Rio do Janeiro, Rio de Janeiro, 2013.

CHOMSKY, N. **Knowledge of language: its nature, origin and use**. London: Praeger Publishers, 1986.

COELHO, I. L. [et al]. **Para conhecer Sociolinguística**. São Paulo: Contexto, 2015.

EDWARDS, J. R. **Multilingualism**. London: Routledge. Penguin Language & Linguistics. 1994.

FARACO, C. A.; VIEIRA, F. E. **Gramáticas brasileiras**: com a palavra, os leitores. São Paulo: Parábola, 2016.

INTUMBO, Incanha. **Estudo comparativo da morfossintaxe do crioulo guineense, do balanta e do português**. Dissertação de Mestrado. Universidade de Coimbra, 2007.

JOUDE, J.. **Em Guiné-Bissau, o sistema educativo precisa em grande parte de ser construído**. Biblioteca Digital da UNESCO. 2016.

LABOV, W. Some sociolinguistic principles. In: Paulston, C. B. & Tucker, G. R. (orgs.). **Sociolinguistics: the essential readings**, 235-250. Oxford: Blackwell, 2003.

_____. **Padrões sociolinguísticos**. Tradução: Marcos Bagno, Maria Marta Pereira Scherre e Caroline Rodrigues Cardoso. São Paulo: Parábola Editorial, 2008.

LIMA-HERNANDES, M. C. Sociolinguística e línguas de herança. In: MOLLICA, M. C. e FERRAREZI JÚNIOR, C. (Orgs). **Sociolinguística, sociolinguísticas**. São Paulo: Contexto, 2016. p.97-110. (2016),

LUCCHESI, D. Crioulística. In: MOLLICA, M. C. e FERRAREZI JÚNIOR, C. (Orgs). **Sociolinguística, sociolinguísticas**. São Paulo: Contexto, 2016. p. 73-85. 2016

MOLLICA, M. C. Fundamentação Teórica: conceituação e delimitação / Relevância das variáveis não linguísticas. In: MOLLICA, M. C.; BRAGA, M. L. (Orgs). **Introdução à Sociolinguística: o tratamento da variação**. São Paulo: Contexto, 2013. p. 09-14; 27-32.

MOTA, M. A.; MIGUEL, M.; MENDES, A. A concordância de P6 em português falado. Os traços pronominais e os traços de concordância. **Papia**, v. 22, n. 1, p. 161-187, 2012.

NAMONE, D.; TIMBANE, A. A.. Consequências do ensino da língua portuguesa no ensino fundamental na Guiné-Bissau 43 anos após a independência. **Mandinga - Revista de Estudos Linguísticos**, [S.l.], v. 1, n. 1, p. 39-57, 2017. ISSN 2526-3455. 2017.

PAIVA, M. C. A variável gênero/sexo. In: MOLLICA, M. C.; BRAGA, M. L. (Orgs). **Introdução à Sociolinguística: o tratamento da variação**. São Paulo: Contexto, 2013. p. 35-42.

PEIXOTO, C. M. M; SOARES, M. E. **Representações sobre o estatuto da língua portuguesa no dizer de estudantes guineenses**. XVII CONGRESSO INTERNACIONAL ASOCIACIÓN DE LINGÜÍSTICA Y FILOLOGÍA DE AMÉRICA LATINA (ALFAL 2014) João Pessoa - Paraíba, Brasil. 2014.

PETTER, M. M. T. Aspectos morfossintáticos comuns ao Português angolano, brasileiro e moçambicano. **Revista PAPIA** 19, p. 201-220. Universidade de São Paulo. 2009.

REVUZ, Christine. A língua estrangeira: entre o desejo de um outro lugar e o risco do exílio. In: SIGNORINI, Inês (Org.). **Língua(gem) e identidade**. São Paulo: Mercado das Letras, p. 213-230, 1998.

SANKOFF, D.; TAGLIAMONTE, S.; SMITH, E. **GoldVarb X – a multivariate analysis application**. 2005. Toronto: Department of Linguistics; Ottawa: Department of Mathematics.

SANTOS, O. F. **Sobre o parâmetro do sujeito nulo**: Interferência do Português Brasileiro no Espanhol falado no Brasil por Nativos residentes no Brasil. Dissertação de Mestrado. Programa de pós-graduação em Letras da UFPE. 146 P. 2013.

SCHERRE, M. M. P. **Reanálise da concordância nominal em português**. Tese de Doutorado em Linguística. Rio de Janeiro, Faculdade de Letras, UFRJ, 1988.

SCHERRE, M. M. P.; NARO, A. J. **Sobre a concordância de número no português falado do Brasil**. In Ruffino, Giovanni (org.) *Dialettologia, geolinguística, sociolinguística*. (Attidel XXI Congresso Internazionale di Linguistica e Filologia Romanza) Centro di Studi Filologici e Linguistici Siciliani, Università di Palermo. Tübingen: Max Niemeyer Verlag, 5:509-523, 1998.

SCHERRE, M. M. P.; NARO, A. J. **Mudança sem mudança**: a concordância de número no português brasileiro. *SCRIPTA*, Belo Horizonte, v. 9, n. 18, p. 107-129, 1º sem. 2006.

SEDRINS, A. P.; SILVA, C. R. T. Padrões de concordância de gênero e número no sintagma nominal em variedades africanas do português. *Leitura. Revista do Programa de Pós-Graduação em Letras (UFAL)*, v. 2, p. 85-105, 2017.

SILVA, C. K. B. da. **Variação da concordância nominal em produção oral e escrita de alunos do ensino fundamental e médio de Belo Jardim-PE**: assimetria entre fala e escrita? Dissertação de Mestrado. Recife-PE: Universidade Federal de Pernambuco. Programa de Pós-graduação em Letras. 2017. 132 p.

SILVA, V. L. P. Relevância das variáveis linguísticas. In: MOLLICA, M. C.; BRAGA, M. L. (Org.). **Introdução à Sociolinguística**: o tratamento da variação. São Paulo: Contexto, p. 67-72. 2013.

TARALLO, F. **A pesquisa sociolinguística**. São Paulo: Ática, 1985.

VIEIRA, S. R.; BRANDÃO, S. F. Tipologia de regras linguísticas e estatuto das variedades/línguas: a concordância em português. *Linguística* / Vol. 30 (2). 81-112 ISSN 1132-0214 impresa ISSN 2079-312. Dezembro, 2014.