



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA
DEPARTAMENTO DE LETRAS E ARTES
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS

LAURA CAMILA DOS SANTOS SANTANA

**O USO DO PRONOME NULO E DO PRONOME PLENO NA POSIÇÃO SUJEITO
NA LÍNGUA FALADA DA COMUNIDADE QUILOMBOLA BAIXA DA LINHA NO
RECÔNCAVO BAIANO**

A presente pesquisa foi realizada com o apoio da Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB).

Feira de Santana-BA

2025

LAURA CAMILA DOS SANTOS SANTANA

**O USO DO PRONOME NULO E DO PRONOME PLENO NA POSIÇÃO SUJEITO
NA LÍNGUA FALADA DA COMUNIDADE QUILOMBOLA BAIXA DA LINHA NO
RECÔNCAVO BAIANO**

Tese apresentada à banca examinadora do Programa de Pós-Graduação em Estudos Linguísticos, da Universidade Estadual de Feira de Santana, como requisito para o título de Doutora em Estudos Linguísticos.

Orientadora: Profa. Dra. Silvana Silva de Farias Araujo.

Feira de Santana-BA

2025

TERMO DE APROVAÇÃO

LAURA CAMILA DOS SANTOS SANTANA

O USO DO PRONOME NULO E DO PRONOME PLENO NA POSIÇÃO SUJEITO NA LÍNGUA FALADA DA COMUNIDADE QUILOMBOLA BAIXA DA LINHA NO RECÔNCAVO BAIANO

TESE DE DOUTORADO

Tese submetida ao Programa de Pós-Graduação em Estudos Linguísticos da Universidade Estadual de Feira de Santana, área de concentração Linguagem e Sociedade, Linha de Pesquisa Variação e mudança linguística no português, como requisito para obtenção do título de Doutora em Estudos Linguísticos.

Aprovada em: 09/10/2025.

BANCA EXAMINADORA

Documento assinado digitalmente



SILVANA SILVA DE FARIAS ARAUJO

Data: 19/03/2026 11:16:51-0300

Verifique em <https://validar.iti.gov.br>

Profa. Dra. Silvana Silva de Farias Araujo
Universidade Estadual de Feira de Santana

Documento assinado digitalmente



NORMA LUCIA FERNANDES DE ALMEIDA

Data: 19/03/2026 17:25:34-0300

Verifique em <https://validar.iti.gov.br>

Profa. Dra. Norma Lucia Fernandes de Almeida
Universidade Estadual de Feira de Santana

Examinador Interno

Documento assinado digitalmente



LILIANE LEMOS SANTANA BARREIROS

Data: 19/03/2026 11:40:22-0300

Verifique em <https://validar.iti.gov.br>

Profa. Dra. Liliane Lemos Santana Barreiros
Universidade Estadual de Feira de Santana

Documento assinado digitalmente



CLAUDIA ROBERTA TAVARES SILVA

Data: 19/03/2026 20:22:17-0300

Verifique em <https://validar.iti.gov.br>

Profa. Dra. Cláudia Roberta Tavares Silva
Universidade Federal Rural de Pernambuco - UFRPE

Documento assinado digitalmente



DANTE EUSTACHIO LUCCHESI RAMACCIOTTI

Data: 20/03/2026 11:58:42-0300

Verifique em <https://validar.iti.gov.br>

Prof. Dr. Dante Eustachio Lucchesi Ramacciotti
Universidade Federal Fluminense - UFF
Examinador Externo

Ficha catalográfica - Biblioteca Central Julieta Carteadó - UEFS

Santana, Laura Camila dos Santos

S223u O uso do pronome nulo e do pronome pleno na posição sujeito na língua falada da comunidade quilombola Baixa da Linha no Recôncavo Baiano / Laura Camila dos Santos Santana. - 2025.
159f.: il.

Orientadora: Silvana Silva de Farias Araújo

Tese (doutorado) - Universidade Estadual de Feira de Santana.
Programa de Pós-Graduação em Estudos Linguísticos, 2025.

1. Sujeito nulo. 2. Português afro-brasileiro. 3. Contato linguístico.
4. Comunidade quilombola. I. Araújo, Silvana Silva de Farias, orient.
II. Universidade Estadual de Feira de Santana. Programa de Pós-Graduação em Estudos Linguísticos. III. Título.

CDU: 806.90

Para (e por) Rosilda e Joselito.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por Sua presença em todas as etapas da minha vida.

Aos meus amados pais, Rosilda e Joselito, que sempre me deram apoio, incentivo e seus braços abertos. Aos meus irmãos e sobrinhos por sempre estarem ao meu lado e acreditarem que esta tarefa poderia ser concluída. A todos os meus familiares e amigos que acreditaram e me fortaleceram durante todo o processo. Sem dúvidas, sem esse alicerce que me motivou, eu não conseguiria chegar até aqui.

É imprescindível lembrar de pessoas queridas e instituições que forneceram orientações, reflexões e sugestões para o desenvolvimento deste trabalho. Portanto, agradeço:

À professora Silvana Silva de Farias Araújo, pois além de ser uma brilhante orientadora e pesquisadora, acreditou em meu potencial para levar este estudo adiante, mesmo em face dos desafios encontrados ao longo do caminho. Seus conselhos e orientações foram valiosos para mim. Agradeço pelas portas que a disciplina no mestrado “Estudo de Variedades do Português” conseguiu abrir, me fascinando para a linha de pesquisa 1, Variação e mudança linguística no português.

Ao professor Dante Lucchesi, cujas contribuições teóricas são referências para a Sociolinguística brasileira, pelas orientações, e materiais que contribuíram de modo preponderante para esta pesquisa, desde a fase de arguição do projeto até a conclusão desta tese.

À professora Norma Lucia Fernandes de Almeida e à professora Cláudia Roberta Tavares Silva não apenas pelas relevantes contribuições para esta pesquisa, mas também por serem, juntamente à Silvana Araújo, inspiração para muitos pesquisadores.

Aos moradores da comunidade Baixa da Linha que me permitiram desfrutar das mais belas e intensas narrativas de vida, em especial a Márcia Cristina e José, duas lideranças fundamentais.

À Jean Marcel Araújo por sua atenção e disponibilidade ao nos auxiliar prontamente na fase burocrática do ser pesquisador.

Aos bolsistas de iniciação científica, Matheus, Sarah e Alice que contribuíram com a transcrição grafemática de entrevistas que fazem parte deste trabalho.

À Damares Oliveira de Souza pelo apoio, incentivo e trocas acadêmicas virtuais ao longo desses anos de pesquisa.

À Jacson Baldoíno Silva pelos diálogos compartilhados e pela motivação que me concedeu.

À Eduardo Patrick Rezende por me estender a mão e demonstrar com serenidade e leveza como o pesquisador pode superar os desafios.

Aos meus queridos alunos do ensino fundamental e médio, que fazem parte do meu laboratório real de pesquisa.

À Universidade Estadual de Feira de Santana (UEFS), pelo seu acolhimento institucional, manifestado em serviços como a disponibilização de livros e acervos históricos da biblioteca, bem como a presteza de todos os servidores.

Ao Programa de Pós-graduação em Estudos Linguísticos (PPGEL), pela exímia manutenção e condução do curso, mesmo diante do cenário abrupto de uma pandemia mundial. Sou grata por todo o corpo docente, técnicos e demais funcionários, por todas as aulas remotas e pelas oportunidades de acessar outros espaços de construção de conhecimentos oportunizados pela tecnologia.

Ao Núcleo de Estudos de Língua Portuguesa (NELP), cujo objetivo é desenvolver na UEFS, e em diálogo com pesquisadores de outras instituições, pesquisas sobre a história da língua portuguesa, não apenas no Brasil, mas também na África.

À Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) pelo financiamento desta pesquisa.

“As primeira pessoas que chegou aqui era tudo do lugar [...] que trabalha Bahia [...] Naquela usina de açúcar que os povo pegava os povo por aí a pulso, trazia pra trabalhar de qualquer jeito sem ganhar dinheiro, sem nada, e tinha que ir trabalhar. Aí eu fui ver que meu avô [...] o pai da minha vó era do mato” (Informante SS03 – *corpus* da pesquisa).

RESUMO

A presente tese analisa o preenchimento da posição de sujeito numa variedade do português a partir de um *corpus* constituído de 909 dados orais coletados na comunidade rural afro-brasileira Baixa da Linha, no Recôncavo Baiano. A amostra de fala é composta por 10 entrevistas realizadas na comunidade, cujos participantes foram estratificados em sexo, faixa etária, escolaridade e estada fora da comunidade. Busca-se, a partir da descrição, identificar quais os contextos linguísticos e socioculturais que favorecem ou inibem a categoria vazia na posição de sujeito, procurando observar o uso variável do preenchimento do sujeito, a fim de investigar a atuação do contato do português com línguas africanas nesta região. Para tanto, o aporte teórico e metodológico utilizado foi a Sociolinguística Variacionista (Weinreich; Labov; Herzog, 2006[1968]; Labov, 1972 [2008]), e os pressupostos teóricos da Sociolinguística Paramétrica (Tarallo, 1987; Tarallo; Kato, 1989; Duarte, 2016, 2019; Martins; Coelho; Cavalcante, 2015). A perda gradual do sujeito nulo no português brasileiro foi constatada por Duarte (1995 [2018]), com a perda do princípio “evite pronomes”. Essa mudança pode ter se consolidado pela redução do paradigma flexional verbal e pela alteração no sistema pronominal com a entrada no sistema das formas *você* e *a gente* (Lopes, 2003; Lopes; Brocardo, 2016; Omena, 1986; 1996; Menon, 1994;1996; Lopes e Cavalcante, 2011). No entanto, convém investigar o comportamento desse fenômeno em outras variedades, como o português afro-brasileiro, de modo a compreender como a dinâmica sócio-histórica desta comunidade afro-brasileira, no Recôncavo Baiano, interferiu no preenchimento ou não da posição sujeito. A partir de estudos já realizados em outras comunidades afro-brasileiras, como o de Lucchesi (2004; 2009); Almeida (2005); Silva (2017), além do trabalho relevante de Duarte (1993; 1995; 2000; 2018, 2019), dentre outros, procuramos discutir aspectos como a história social da comunidade Baixa da Linha, no município de Cruz das Almas-BA; a influência exercida, por meio do contato com outros povos e o efeito da Transmissão Linguística Irregular (Lucchesi; Baxter, 2009) como um fator sócio-histórico decisivo para a realização ou não realização do sujeito pronominal no português afro-brasileiro; identificar os contextos de maior favorecimento do sujeito nulo, bem como seus possíveis contextos de resistência; prever possíveis efeitos provocados por essa mudança no português afro-brasileiro. A análise teve como base o feixe de propriedades linguísticas associadas ao Parâmetro do Sujeito Nulo, sobretudo os de Lucchesi (2009) e Duarte (1995, 2012 e 2018). Após a identificação das ocorrências, as mesmas foram codificadas e submetidas ao programa Goldvarb X, que apresentou as frequências e pesos relativos para as variáveis investigadas. O programa selecionou os seguintes grupos de fatores como relevantes: tipo de oração, menção anterior do sujeito, localização em relação ao antecedente e estrutura sintática da oração. Entretanto, apesar de nenhum fator social haver sido selecionado pelo programa, nota-se uma nítida preferência dos falantes mais velhos pelo sujeito foneticamente preenchido, o que confirma a hipótese de que o contato entre línguas no Brasil alterou a estruturação das variedades populares do PB devido ao processo de Transmissão Linguística Irregular em sua formação, indicando um processo adiantado de mudança na gramática da comunidade com um traço negativo ao parâmetro *pro-drop* muito anterior às mudanças que seriam observadas no PB culto (Tarallo, 1993; Duarte, 1993, 1995, 2018).

PALAVRAS-CHAVE: sujeito nulo; português afro-brasileiro; contato linguístico; comunidade quilombola.

ABSTRACT

This thesis analyzes the filling of the subject position in a variety of Portuguese based on a *corpus* of 909 oral data collected in the rural Afro-Brazilian community Baixa da Linha, in the Bahian Recôncavo region. The speech sample consists of 11 interviews conducted in the community, whose participants were stratified by gender, age, education level, and residence outside the community. Based on the descriptions, the aim is to identify which linguistic and social contexts favor or inhibit the empty category in the subject position, seeking to observe the variable use of the subject filling, in order to investigate the impact of Portuguese's contact with African languages in this region. To this end, the theoretical and methodological framework used was Variationist Sociolinguistics (Weinreich; Labov; Herzog, 2006[1968]; Labov, 1972 [2008]), and the theoretical assumptions of Parametric Sociolinguistics (Tarallo, 1987; Tarallo; Kato, 1989; Duarte, 2016, 2019; Martins; Coelho; Cavalcante, 2015). The gradual loss of the null subject in Brazilian Portuguese was observed by Duarte (1995 [2018]), with the loss of the “avoid pronoun” principle. This change may have been consolidated by the reduction of the verbal inflectional paradigm and by the change in the pronominal system with the entry into the system of the forms *você* and *a gente* (Lopes, 2003; Lopes; Brocardo, 2016; Omena, 1986; 1996; Menon, 1994; 1996; Lopes and Cavalcante, 2011). However, it is worth investigating the behavior of this phenomenon in other varieties, such as Afro-Brazilian Portuguese, in order to understand how the socio-historical dynamics of this Afro-Brazilian community, in Recôncavo Baiano, interfered with the filling or not of the subject position. Based on studies already carried out in other Afro-Brazilian communities, such as Lucchesi (2004; 2009); Almeida (2005); Silva (2017), in addition to the relevant work of Duarte (1993; 1995; 2000; 2018, 2019), among others, we seek to discuss aspects such as the social history of the Baixa da Linha community, in the municipality of Cruz das Almas-BA; the influence exerted through contact with other peoples and the effect of Irregular Linguistic Transmission (Lucchesi; Baxter, 2009) as a decisive socio-historical factor for the realization or non-realization of the pronominal subject in Afro-Brazilian Portuguese; identify the contexts of greatest favoritism of the null subject, as well as its possible contexts of resistance; predict possible effects caused by this change in Afro-Brazilian Portuguese. The analysis was based on the set of linguistic properties associated with the Null Subject Parameter, particularly those of Lucchesi (2009) and Duarte (1995, 2012, and 2018). After identifying the occurrences, they were coded and submitted to the Goldvarb X program, which presented the frequencies and relative weights for the investigated variables. The program selected the following groups of factors as relevant: type of sentence, previous mention of the subject, location in relation to the antecedent, and syntactic structure of the sentence. However, despite no social factor having been selected by the program, a clear preference is noted among older speakers for the phonetically filled subject, which confirms the hypothesis that language contact in Brazil altered the structuring of popular varieties of Brazilian Portuguese due to the process of Irregular Linguistic Transmission in their formation, indicating an advanced process of change in the community's grammar with a negative trait to the pro-drop parameter much earlier than the changes that would be observed in standard Brazilian Portuguese (Tarallo, 1993; Duarte, 1993, 1995, 2018).

KEYWORDS: null subject; afro-Brazilian Portuguese; language contact; quilombola community.

LISTA DE FIGURAS

Figura 1 -	Mapa do Recôncavo da Bahia de Todos os Santos (1836)	23
Figura 2 -	Municípios que compõem o Território de Identidade do Recôncavo Baiano (2015)	29
Figura 3 -	Fragmentos de cerâmica, líticos lascados e urnas funerárias	32
Figura 4 -	Vista área da Universidade Federal do Recôncavo da Bahia (UFRB)	36
Figura 5 -	Apresentação cultural de samba de roda na comunidade Baixa da Linha	37
Figura 6	Antiga Estação de trem de Cruz das Almas	38
Figura 7 -	Vista da comunidade Baixa da Linha	39
Figura 8 -	Associação comunitária quilombola da Baixa da Linha	40

LISTA DE QUADROS

Quadro 1 - Comunidades rurais afro-brasileiras remanescentes de quilombos no Recôncavo Baiano	30
Quadro 2 - Ilustração da mudança do PSN durante a aquisição linguística na situação afro-brasileira	50
Quadro 3 - Frequência de sujeitos nulos no PE	53
Quadro 4 - Evolução nos paradigmas flexionais do PB	55
Quadro 5 - Retenção pronominal	58
Quadro 6 - Demografia histórica da sociedade brasileira	82
Quadro 7- Perfil dos participantes da pesquisa	89

LISTA DE GRÁFICOS

Gráfico 1 -	Categorias sociais no Recôncavo no início do século XVII	26
Gráfico 2 -	Sujeito pronominal expesso e sujeito nulo	105
Gráfico 3 -	Sujeito pronominal expesso e sujeito nulo sem orações coordenadas	106

LISTA DE TABELAS

Tabela 1 -	Distribuição de comunidades quilombolas por região	30
Tabela 2 -	Pronomes nominativos no português europeu (PE)	44
Tabela 3 -	Quadro pronominal do português brasileiro (PB)	45
Tabela 4 -	Quadro pronominal do português afro-brasileiro (PAF)	46
Tabela 5 -	Hierarquia Referencial	56
Tabela 6-	Grupos de fatores selecionados como significantes para a ocorrência do preenchimento do sujeito segundo o programa	104
Tabela 7-	Percentual de sujeitos expressos <i>versus</i> sujeito nulo	107
Tabela 8-	Sujeitos nulos com relação às pessoas gramaticais	111
Tabela 9-	Sujeitos nulos relacionados às pessoas gramaticais separadamente	112
Tabela 10-	O sujeito expresse e a Desinência gramatical do verbo	116
Tabela 11-	Sujeito nulo e caracterização semântica do sujeito	119
Tabela 12-	Sujeito expresse em relação ao tipo de oração	121
Tabela 13-	Sujeito expresse e menção anterior do sujeito	123
Tabela 14-	Sujeito nulo e localização em relação ao antecedente	128
Tabela 15-	Sujeito nulo e estrutura sintática da oração	130
Tabela 16-	Sujeito nulo e sexo dos informantes	132
Tabela 17-	Sujeito realizado e faixa etária dos informantes	133

LISTA DE SÍMBOLOS E ABREVIATURAS

AGR	Concordância
CRQ	Comunidade Remanescente de Quilombo
CV	Categoria vazia
DID	Diálogo entre Informante e Documentador
LL	Língua Lexificadora
PAF	Português Afro-brasileiro
PB	Português Brasileiro
PPB	Português Popular do Brasil
PE	Português Europeu
PR	Peso Relativo
PSN	Parâmetro do Sujeito Nulo
SN	Sintagma nominal
SPD	Sociedade Protetora dos Desvalidos
TPP	Teoria de Princípios e Parâmetros
TLI	Transmissão Linguística Irregular

SUMÁRIO

1	INTRODUÇÃO	17
2	ASPECTOS SÓCIO-HISTÓRICOS DO RECÔNCAVO BAIANO	21
2.1	Comunidades Quilombolas no Recôncavo Baiano	27
2.1.1	A Comunidade quilombola Baixa da Linha	30
3	A REALIZAÇÃO DO SUJEITO PRONOMINAL EM VARIEDADES DO PORTUGUÊS	40
3.1	Pronomes pessoais: algumas discussões	38
3.1.1	Quadro pronominal em variedades do português: europeu, brasileiro e afro-brasileiro	41
3.2	O Parâmetro do Sujeito Nulo	41
3.3	O Parâmetro do Sujeito Nulo em Variedades do Português	52
4	PRESSUPOSTOS TEÓRICOS	63
4.1	A Sociolinguística Variacionista	63
4.1.1	As variáveis linguísticas dentro do sistema	67
4.1.2	As cinco questões da mudança linguística	69
4.2	O Contato linguístico	73
4.2.1	A Transmissão Linguística Irregular (TLI)	75
4.2.2	A polarização sociolinguística do Brasil	78
4.3	O Contexto Sócio-Histórico Da Bahia	81
4.3.1	A pressão demográfica e a escolarização no Brasil	81
4.3.1.2	<i>O processo de escolarização na Bahia</i>	84
5	ASPECTOS METODOLÓGICOS	86
5.1	A Constituição das amostras de fala	86
5.1.1	Coleta de dados orais	87
5.1.2	O perfil dos informantes	88
5.1.3	A natureza das entrevistas	90
5.2	O envelope da variação	91
5.2.1	A variável dependente	91
5.2.1.1	<i>Crítérios para inclusão e exclusão de dados</i>	91
5.2.1.2	<i>As variáveis linguísticas</i>	92
5.2.1.2.1	Pessoa do sujeito	92
5.2.1.2.2	Desinência gramatical do verbo	94

5.2.1.2.3	Caracterização semântica do sujeito	95
5.2.1.2.4	Tipo de frase	96
5.2.1.2.5	Paralelismo formal	96
5.2.1.2.6	Localização em relação ao antecedente e as condições de referencialidade	97
5.2.1.2.7	Presença de constituinte à esquerda do verbo na oração	99
5.2.1.3	<i>As variáveis sociais</i>	100
5.2.1.3.1	<i>Sexo</i>	100
5.2.1.3.2	<i>Faixa etária</i>	101
5.2.1.3.3	<i>Escolaridade</i>	101
5.3	Processamento Quantitativo dos dados	102
6	ANÁLISE VARIACIONISTA DA REALIZAÇÃO DO SUJEITO PRONOMINAL NA COMUNIDADE BAIXA DA LINHA-BA	103
6.1	A variável dependente	105
6.2	As variáveis linguísticas	110
6.2.1	Pessoa do sujeito	110
6.2.2	Desinência gramatical do verbo	
6.2.3	Caracterização semântica do sujeito	118
6.2.4	O tipo de oração	120
6.2.5	Menção anterior ao sujeito	122
6.2.6	Localização em relação ao antecedente	127
6.2.7	Estrutura sintática da oração	130
6.3	As variáveis sociais	131
6.3.1	Sexo	131
6.3.2	Faixa etária	133
7	CONSIDERAÇÕES FINAIS	136
	REFERÊNCIAS	140
	APÊNDICES	150
	ANEXOS	151

1 INTRODUÇÃO

A indagação sobre o sujeito nulo no Português Brasileiro (PB) é um assunto de intensos debates. No campo da investigação linguística, o estudo do comportamento desse fenômeno tem levado muitos pesquisadores a defenderem uma reorganização pela qual vem atravessando a gramática do PB (Duarte, 1995, 2000; Figueiredo Silva, 1996; Kato, 2000; Galves, 2001; Silva, 2004, 2017). Apesar desses esforços, no entanto, ainda há questões para se responder (Almeida; Carneiro, 2009; Almeida; Pitombo, 2019), principalmente em relação ao estatuto do PB no Parâmetro do Sujeito Nulo (PSN). O desenvolvimento desta pesquisa partiu inicialmente da relevância social da comunidade rural Baixa da Linha não apenas para a região do Recôncavo Baiano, mas para todo o contexto brasileiro. Apesar de não termos indícios comprobatórios da permanência de quilombos prototípicos nesta região durante a colonização do Brasil, a forte presença de afrodescendentes na comunidade revela as marcas da exclusão social imposta a esses indivíduos, resistência à escravização e a atual luta pelo seu território.

Este estudo se propõe a romper com a visão comparatista desenvolvida em muitas pesquisas linguísticas, como Almeida (2009), nas quais o objetivo maior era o de estudar o PB à luz do Português Europeu (PE), tomando o primeiro como uma espécie de deturpação do segundo. Sendo assim, o desenvolvimento desta pesquisa tem por motivação a urgência de se estudar a fala das comunidades rurais afro-brasileiras para a compreensão da formação histórica da realidade linguística brasileira *per se*, pois as variedades populares e rurais do PB refletem processos de variação e mudança resultantes do contato do português com as línguas dos africanos escravizados no Brasil, processo esse que marca nossa realidade linguística de forma única. Os dados de fala coletados nesta pesquisa poderão, assim, produzir evidências empíricas valiosas acerca da relevância do contato entre línguas na história sociolinguística do Brasil.

Desse modo, convém investigar como a dinâmica sócio-histórica da comunidade rural afro-brasileira Baixa da Linha, no Recôncavo Baiano, interferiu para o preenchimento ou não da categoria de sujeito no âmbito do PSN, buscando ali indícios que possam servir para explicar ou, ao menos, refinar as explicações sobre o quadro brasileiro como um todo. Essa questão de investigação diz respeito ao modo como essa variedade, que se formou a partir do contato linguístico intenso entre a língua portuguesa e línguas africanas, se comporta em relação ao PSN. Outra questão neste estudo refere-se a investigar se o Português Afro-brasileiro (PAF)

falado nesta comunidade exibe uma tendência oposta de mudança em relação à norma brasileira, no sentido de aumento da frequência de sujeitos nulos, ou se exibe um estágio mais avançado de perda de licenciamento do sujeito referencial nulo, aproximando-se mais das línguas crioulas de base lexical portuguesa da África (Lucchesi, 1993, 2009).

O PB, nesse sentido, possui características que o diferenciam do PE (Duarte, 1993, 1995, 2018, 2019, 2020; Cyrino, Moraes, 2018; Roberts, Kato, 2018[1993]; Galves, Kato, Roberts, 2019). Uma das razões apontadas para tal distinção diz respeito aos intensos contatos entre línguas ocorrido ao longo do processo de exploração do território brasileiro, desencadeado pela colonização do Brasil (Almeida, 2005; Lobo, 2018; Tarallo, 2018[1993]; Roberts, 2019; Lucchesi, 2019). Dentre os estudos comparativos entre o PE e o PB, o PSN tem sido um dos mais estudados, recebendo assim mais refinamentos e contribuições (Duarte, 2018). O estudo de Duarte procurou demonstrar que uma abordagem variacionista pode contribuir para o tratamento formal das questões linguísticas (Duarte, 1993, 1995, 2019). Muito embora os estudos de Duarte sejam de grande relevância para a compreensão de características que individualizam o PB, à luz da Teoria de Princípios e Parâmetros (TPP) (Chomsky, 1981, 1982, 1986), analisar esse mesmo parâmetro em outras variedades do português é fundamental para a compreensão dos processos de variação e mudança pelos quais passa a língua, haja vista a formação bipolarizada da língua portuguesa no Brasil (Lucchesi, 2001, 2002, 2006, 2015).

A presente pesquisa de doutorado tem como macro-objetivo analisar o preenchimento do sujeito a partir de um *corpus* constituído por 12 entrevistas coletadas na comunidade rural afro-brasileira Baixa da Linha, localizada no Recôncavo Baiano. Para tanto, esta pesquisa tem como aporte teórico a Sociolinguística Variacionista (Weinreich; Labov; Herzog, 1968[2006]; Labov, 1972 [2008]), a qual fornece o suporte de análise do panorama sociolinguístico do PB a partir das implicações do contato entre línguas na sua formação sócio-histórica. O fenômeno de investigação é a realização do sujeito pronominal, cujo estudo em comunidades afro-brasileiras suscitou questionamentos, como os descritos por Lucchesi (2004 [2009]), Almeida (2005), Silva (2017) e Silva (2023), em comunidades quilombolas da região Nordeste do Brasil.

A questão de investigação deste estudo é identificar, nos moldes de Lucchesi (2004 [2009]) e Almeida (2005), como o português afro-brasileiro que se formou a partir do contato linguístico se comporta em relação ao PSN. A partir da descrição, buscaremos identificar quais os contextos linguísticos e sociais que favorecem ou inibem a categoria vazia na posição de sujeito, procurando observar o uso variável do preenchimento do sujeito, a fim de investigar a atuação do contato do português com línguas africanas na comunidade rural Baixa da Linha.

Para compreender os contextos que favorecem o preenchimento do sujeito, serão analisados, como já exposto, dados de fala de uma comunidade rural, remanescente de quilombo, a comunidade Baixa da Linha, em Cruz das Almas-BA. Foram gravadas doze (12) entrevistas do tipo diálogo entre informante e documentador (DID) com moradores dessa comunidade. Os participantes foram selecionados segundo os critérios: faixa etária, sexo, nascimento na localidade e escolaridade. Após as gravações, as entrevistas foram transcritas e os dados codificados segundo as variáveis linguísticas e sociais. A pesquisa contou também com o suporte do Programa GoldVarb X (Sankoff; Tagliamonte; Smith, 2005) para as análises quantitativas e estatísticas. Este trabalho é, portanto, de base empírica e seguirá os pressupostos da análise variacionista quantitativa desenvolvida por Labov (1972) e da pesquisa sociolinguística em sentido amplo (Tarallo, 1985; Mollica, 1992). Neste estudo também utilizaremos parte do saber teórico referente ao PSN desenvolvido em outros quadros, como o gerativismo, pois o julgamos fundamental para formulação de hipóteses e levantamento do grupo de fatores linguísticos/estruturais que condicionam o fenômeno em questão. Assim, tentaremos conciliar¹ duas abordagens epistemológicas: uma empirista (língua-E) e a outra, racionalista (língua- I).

A partir das pesquisas de Duarte (1993, 1995, 2018, 2019), nota-se a perda gradual do sujeito nulo no português brasileiro. Essa mudança parece, de acordo com a autora, ter se consolidado pela redução do paradigma de flexões verbais em função da alteração no sistema pronominal do português com a entrada das formas *você* e *a gente* (Lopes, 2003; Lopes e Brocardo, 2016). Entretanto, para além da norma culta, uma vez que a polarização sociolinguística marca a história do Brasil (Lucchesi, 1994, 1998, 2001, 2002, 2006, 2015), torna-se relevante compreender a manifestação desse parâmetro no português popular brasileiro (PPB), mais especificamente no PAF.

A hipótese adotada nesta pesquisa é a de que o contato do português, sobretudo com as línguas africanas, afetou a estruturação das variedades linguísticas do PB, sobretudo as variedades populares. Desse modo, são as comunidades rurais afro-brasileiras que irão demonstrar os reflexos mais salientes dessa influência. Assim, acreditamos que na comunidade estudada, devido às implicações do contato linguístico, especificamente ao processo de Transmissão Linguística Irregular (TLI) (Lucchesi; Baxter, 2009), a variante mais frequente

¹ Esta união entre a Teoria de Princípios de Parâmetros e a Sociolinguística, um “casamento herético”, foi denominada por Tarallo e Kato (1989) de Sociolinguística Paramétrica. A variação e a mudança linguística são percebidas a partir da perspectiva interlinguística e intralinguística. Dentre as pesquisas que se desenvolveram a partir dessa corrente teórica estão os estudos de Martins, Coelho e Cavalcante (2015).

seria o sujeito preenchido, pois a erosão da morfologia flexional do verbo teria reduzido em muito a possibilidade do sujeito pronominal nulo, de forma semelhante ao processo exibido na variedade guineense do português (Malta e Silva, 2025). Malta (2019), ao analisar dados escritos das variedades africanas do português de Angola, Guiné-Bissau e São Tomé e Príncipe, observa que a concordância verbal apresenta comportamento variável e que a distribuição de sujeitos nulos e plenos não é homogênea entre essas variedades. Enquanto Angola e São Tomé e Príncipe exibem ocorrências relevantes de sujeito nulo, aproximando-se de um comportamento intermediário entre o PB e o PE, o português da Guiné-Bissau apresenta maior incidência de sujeitos plenos, possivelmente em função do intenso contato com línguas crioulas.

Com base na visão de uma formação polarizada do PB desenvolvida por Lucchesi (1994, 2001, 2002, 2006), pode-se vislumbrar um desenvolvimento paralelo, e possivelmente convergente, em que na norma culta estaria em curso uma diminuição na frequência de uso do sujeito nulo desencadeada por um processo de reorganização na pauta dos pronomes sujeito, ao passo que, nas diversas variedades populares, essa propriedade já teria sido restringida em graus diferenciados, devido ao nível da erosão de sua morfologia flexional, determinado pelo processo de TLI em sua formação. Assim, os padrões de uso culto do PB estariam se aproximando do comportamento já estabilizado em suas variedades populares; o que poderia explicar as tendências convergentes de mudança. A posição adotada neste estudo é que a TLI se constitui em um parâmetro sócio-histórico (Lucchesi; Baxter, 2009; Mufwene, 2019; Silva; Almeida, 2023) que proporciona maiores explicações sobre a variação em determinados traços linguísticos, especialmente aqueles diretamente implicados no processo de aquisição, como as marcações paramétricas.

Os capítulos da presente tese estão organizados da seguinte forma: na seção 2 apresentamos alguns aspectos sócio-históricos do Recôncavo Baiano, principalmente a formação de comunidades quilombolas, em especial, à da Baixa da Linha; em seguida, na seção 3, discutimos sobre o Parâmetro do Sujeito Nulo a partir de uma revisão desse conceito e sua aplicação no PB bem como às suas variedades; na seção 4 exibimos os pressupostos teóricos que sustentam este estudo como a Sociolinguística Variacionista, o contato linguístico e os efeitos da Transmissão Linguística Irregular, principalmente nas comunidades remanescentes de quilombo; em seguida, na seção 5, apresentamos os aspectos metodológicos da pesquisa, desde a constituição das amostras de fala até o processamento quantitativo dos dados, descrevendo o percurso metodológico adotado neste estudo; na seção 6, por sua vez,

apresentamos os resultados da análise variacionista da realização do sujeito pronominal na comunidade Baixa da Linha; por fim, apresentamos algumas considerações.

2 ASPECTOS SÓCIO-HISTÓRICOS DO RECÔNCAVO BAIANO

“[...] a língua portuguesa era adquirida nas situações as mais precárias pelos escravos, que muitas vezes preferiam se comunicar entre si, usando uma língua franca africana. Sob essas ásperas condições, a língua portuguesa se foi disseminando entre a população pobre, de origem predominantemente indígena e africana, nos três primeiros séculos da história do Brasil (Lucchesi, 2001, p. 102).

Nesta seção, são apresentados alguns aspectos sócio-históricos do Recôncavo Baiano, considerados fundamentais para a compreensão da formação social e cultural da região. Em seguida, discute-se a distribuição e a caracterização das comunidades quilombolas no Recôncavo Baiano, com ênfase na comunidade da Baixa da Linha.

Segundo Leão (1989), no processo de povoamento nos primeiros séculos da colonização do Estado da Bahia, houve uma concentração da população ao longo da planície costeira. A maior predominância nesta região, até o século XVI, era composta por comunidades de origens indígenas, como as vilas de Caiuru, Camamu e Boipeba. Na década de 1930, Gilberto Freyre (1997) definiu a sociedade que se desenvolveu nas regiões litorâneas do Nordeste brasileiro entre meados do século XVI e o final do século XIX, período posterior ao estudo por Leão, com três palavras: monocultura, latifúndio e escravidão. Esse entendimento é consenso entre historiadores como Maria Yedda Linhares que denomina este período da história do Brasil de “visão plantacionista” (Linhares, 2008).

O estudo realizado por Araujo e Araujo (2023), em complemento a esse panorama, parte do entendimento de que a presença massiva de africanos escravizados na Bahia desempenhou um papel que não pode ser minimizado nos estudos históricos e descritivos do PB, uma vez que esses povos permaneceram na Bahia por mais de três séculos, em número muitas vezes superior ao dos colonos europeus e seus descendentes.

O Recôncavo Baiano, por diversas características geográficas, é uma das áreas produtivas mais antigas do Brasil, remontando a época da distribuição das capitanias hereditárias. Na figura 1, a seguir, é apresentado o primeiro mapa do Recôncavo Baiano.

Figura 1 - Mapa do Recôncavo da Bahia de Todos os Santos (1836)



Fonte: Biblioteca Nacional do Brasil

Publicado em Londres, no ano de 1836, o mapa elaborado por von Busch e Weyl exibiu uma vista panorâmica da cidade de Salvador, contendo rios, vilas, aldeias, capelas, fortes e caminhos. Segundo Araújo (2017), o Recôncavo Baiano, além da costa, também foi preferido para a ocupação dos portugueses tanto pela sua localização de fácil acesso por vias marítima e fluvial quanto pela riqueza de suas terras que viabilizavam a exploração agrícola.

A lavoura açucareira tornava esta região uma das mais produtivas da colônia portuguesa, pois proporcionava um doce estado de riquezas aos senhores de engenho, ao passo que causava tortura e desolação aos cativos. Sousa (2009) argumenta que a população que habitava ao redor da Baía de Todos os Santos no ano de 1700 não era composta apenas de escravizados e senhores de engenho, havia também “pessoas livres, plantadores de fumo, lavradores de mandioca, dentre os quais alguns eram agregados dos engenhos e fazendas de cana, vivendo da agricultura de subsistência” (Souza, 2009, p. 54). Esse contexto de não isolamento dos povos africanos e afrodescendentes converge com os estudos de Lucchesi (2019) sobre os motivos que levaram o português brasileiro a não se criouliizar, pois, enquanto dezenas de línguas crioulas se formaram no Caribe, no Brasil não houve um processo duradouro e representativo de criouliização do português.

Lucchesi (2019) atribui à estrutura socioeconômica açucareira do Nordeste o fato de não ter havido criouliização, porque, ao contrário do que se vê nas *plantations* do Caribe, a exemplo do Haiti – a língua foi totalmente criouliizada, devido ao percentual de franceses ser inferior a

10% enquanto que a maior parcela da população estava concentrada em grandes propriedades. Já no contexto brasileiro, “nos engenhos de açúcar e fazendas de café [...] as crianças pretas passeiam em total liberdade, participando das brincadeiras das crianças brancas e das carícias de todas as mulheres de casa (Mattoso, 2003, p. 128). Esses condicionamentos socioeconômicos foram fundamentais para o português não haver se criouloizado no Brasil.

Embora homens e mulheres pobres, muitas vezes libertos e mestiços, sobrevivessem da pesca de peixes e mariscos, além do cultivo de mandioca, milho e feijão, foi a *plantation* açucareira que se consolidou como o principal vetor da economia do Recôncavo. A partir do final do século XVII, segundo Miller (1997), a produção açucareira tornou-se um dos investimentos mais lucrativos das Américas, sendo implantado em colônias inglesas, holandesas e francesas do Caribe como um modelo agroexportador; por esse motivo, a coroa portuguesa passou a estimular essa produção em seus territórios. Os estudos de Schwartz (2001) revelaram, nesse sentido, que a produção de açúcar no Recôncavo remonta a meados do século XVI; e, segundo ele, no início do século XIX, os engenhos da região forneciam talvez a terça parte de todo o açúcar exportado pelo Brasil. Um grande destaque para a produção era a freguesia de Santiago do Iguape, onde os primeiros engenhos foram construídos no final do século XVI (Schwartz, 2001), já sendo conhecida como uma das freguesias açucareiras mais ricas e mais produtivas da Bahia (Vilhena, 1969).

O estudo de Araujo e Araujo (2009, p. 109) revelou que a área canavieira estava localizada no Recôncavo e abrangia as localidades de São Francisco do Conde, Santo Amaro, Cachoeira e Salvador. Os autores revelaram ainda que, para a funcionalidade da *plantation*, havia ainda uma incipiente atividade de criação de gado e plantação de fumo, já que estes serviam de moeda de troca na compra de escravizados. Nota-se, assim, como o cultivo de cana-de-açúcar, no projeto colonizador português, tornou-se uma grande oportunidade de obtenção de lucros².

Ao iniciar o século XVII, o Recôncavo já se achava povoado por mais de 90.000 pessoas, das quais 50.000 eram escravizados (Azevedo, 1957). Assim, conforme Araújo (1970), no final do século XVII surgiram três vilas integradas ao Recôncavo: Nossa Senhora da Ajuda de Jaguaripe, Nossa Senhora do Rosário de Cachoeira e São Francisco da Barra de Sergipe do Conde. Já no início do século XVIII, foram criadas as vilas de Santo Amaro da Purificação, de São Bartolomeu de Maragogipe, de Abrantes do Espírito Santo e da Abadia. Araujo e Araujo

² Araujo e Araujo (2009) afirmam que a cana-de-açúcar era praticamente escassa na Europa, o Recôncavo Baiano mostrou-se particularmente favorável a esse projeto agroexportador, tendo-se tornado, ao lado de Pernambuco, o grande centro produtor de açúcar da Colônia.

(2009, p. 112) destacam que “nos termos³ de Salvador, São Francisco do Conde e Santo Amaro, concentrava-se 55% dos engenhos localizados no Recôncavo e, conseqüentemente, um grande contingente de escravos”. Embora seja difícil precisar o número de africanos escravizados trazidos para o Recôncavo, além de suas origens e línguas maternas, Araujo e Araujo (2009, p. 114) afirmam, com base nas duas avaliações⁴ realizadas no “Engenho freguesia”, “que era bastante diversificada a situação de contato entre línguas”.

Para Stuart Schwartz (2001), no período colonial, “falar da Bahia era falar do Recôncavo, e este foi sempre sinônimo de engenhos, açúcar e escravos”. Vale ressaltar, no entanto, que esta visão parece ter predominado até meados do século XIX, uma vez que a agricultura de exportação (açúcar, fumo e café) gerava excelentes lucros para lavradores, comerciantes e senhores de engenho, ao passo que os escravizados precisavam garantir sua subsistência. A demanda por “gêneros de primeira necessidade”, expressão comum entre os séculos XVIII e XIX, ajudou a moldar a paisagem social e econômica do Recôncavo rural.

A partir dos estudos de Barickman (2003), é possível perceber que, no tocante à alimentação dos escravizados no Recôncavo, o cenário era de precariedade, “pois eles se alimentavam de qualquer coisa comestível que estivesse ao seu alcance” (Barickman, 2003, p. 92), sendo os dois componentes principais a carne seca e a farinha de mandioca⁵. A Abolição e a República, por sua vez, não parece ter mudado radicalmente tal contexto, já que entre 1940 e 1970, os moradores pobres dos distritos rurais que trabalhavam nos canaviais e nos sítios de fumo do Recôncavo continuavam a depender, para sua sobrevivência, da farinha.

Ao examinar a demanda urbana da farinha de mandioca em Salvador, Barickman (2003) aponta a chegada de centenas de barcos ao porto de Salvador, “vindos de vários pontos do Recôncavo juntamente com os navios maiores do comércio costeiro e transatlântico colocam em terra até 2.200 marinheiros por dia. Os navios que vinham da África traziam não apenas marinheiros, mas também escravos” (p. 99). O porto de Salvador, um dos mais importantes

³ Segundo Prado Jr. (1999), Portugal estendeu no Brasil tanto a sua forma de organização do espaço quanto seu sistema político. Dividida em comarcas, a capitania (chamada então de província) era a unidade administrativa central. Essas eram compostas de termos, divididos em freguesias.

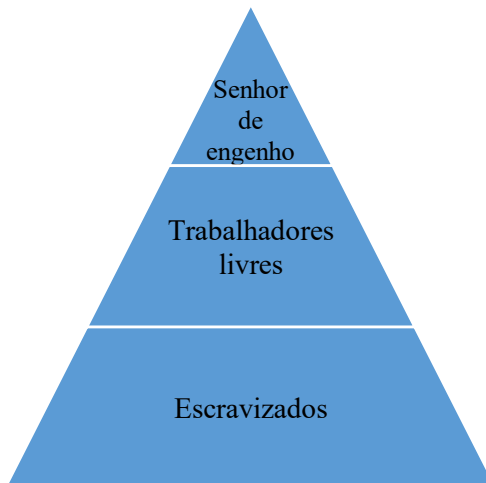
⁴ Pinho [(1982, *apud* Araujo e Araújo (2009, p. 114)] registram duas avaliações de escravizados realizadas no Engenho freguesia localizado na Freguesia de Nossa Senhora da Piedade em Matoim. A primeira foi realizada perante o Juiz de Órfãos Joaquim Anselmo Alves Branco Munis Barreto em 15 de março de 1811. Arrolaram-se 82 escravos, assim etnicamente distribuídos: 13 crioulos, 28 pardos, 16 da costa, 07 jejes, 04 mulatos, 03 nagôs, 02 angolas e 01 mina. Em 14 de fevereiro de 1832, realizou-se a segunda com evidente redução do número de escravos para trinta e quatro: 15 crioulos, 13 pardos, 09 cabras, 04 ussás, 04 mulatos, 04 jejes, 02 mestiços, 01 mina e 01 mendobi.

⁵ Na década de 1870, o viajante alemão Julius Nacher constatou que os escravos no Engenho Subaé, em Santo Amaro, recebiam como alimentação diária apenas um pedaço de carne seca e 1,5 ou dois litros de farinha.

para o tráfico negreiro, abrigava os cativos africanos recém-chegados até que fossem vendidos para engenhos e fazendas no interior ou deslocados para outras partes do Brasil. A permanência de marinheiros e africanos na cidade delineia não apenas o comércio de Salvador, como também modelou a variedade do português que aqui se desenvolveu. Araujo e Araujo (2009, p. 98) afirmam que não há precisão quanto ao ano em que os primeiros africanos foram trazidos para o Brasil, mas apresentam o ano de 1549 como o momento em que o tráfico negreiro foi oficializado. Sendo assim, desse período em diante, mesmo após a Independência do Brasil, em 1822, o intercâmbio, quase sempre forçado, entre África e Brasil esteve na base da formação da nação brasileira.

Araujo e Araujo (2009, p. 114) destacam ainda que os escravizados, tanto africanos quanto nascidos no Brasil, “integravam a estrutura da sociedade do Recôncavo Baiano, organizada sob a ótica das relações senhor-escravo”. No modelo proposto por Mattoso (1978), podiam-se identificar três grandes categorias sociais no início do século XVII, conforme gráfico 1 abaixo:

Gráfico 1 – Categorias sociais no Recôncavo Baiano no início do século XVII



Fonte: Adaptado de Mattoso (1978)

Assim, a sociedade era composta pelo senhor de engenho e sua família, no topo da pirâmide social; em seguida, trabalhadores especializados livres, cuja maioria era de origem lusitana; e, por fim, os escravizados, os quais eram classificados como: qualificados, domésticos e os de oito (em maior número). Contudo, no século XIX, acrescenta-se à diversificação da hierarquia social do Recôncavo a presença de outros sujeitos, uma população de indivíduos (homens brancos livres e pobres, negros forros, libertos, “fugidos” e mestiços) e famílias que

se passaram a constituir uma classe camponesa (Schwartz, 2001) ou uma parte acessória da força de trabalho (Kowarick, 1987). A partir dessa dinâmica composta por vários atores sociais, Araujo e Araújo afirmam que:

Tal organização social desautoriza quaisquer afirmações genéricas acerca de uma realidade linguística una e indivisível para o PB, bem como para a região do Recôncavo da Bahia, para a qual se pode veicular a princípio uma situação linguística polarizada: de um lado, aqueles que se encontravam no topo da pirâmide com acesso à educação e, do outro, uma grande maioria da população alijada da educação, visto que não tinha como custear os estudos no Colégio dos Jesuítas e sua conclusão na Europa (Araujo e Araújo, 2009, p. 116).

A reconstrução da história do PB, diante desse cenário de extrema complexidade constatado pelos estudos historiográficos, faz parte dos anseios de muitos linguistas, como Mattos e Silva. Embora alguns linguistas enxergassem o PB como uma realidade homogeneizável, dos quais podemos citar Serafim da Silva Neto (1950, 1960), Sílvio Elia (1940, 1961, 1979), G. Chaves de Melo (1972), Mattos e Silva defende, partindo de Houaiss (1985), Mussa (1991), Tarallo (1995) dentre outros, a heterogeneidade sincrônica (e diacrônica) do PB. É nesse sentido que o estudo do vernáculo coletado em comunidades afro-brasileiras pode revelar traços definidores do percurso histórico da língua falada no Brasil que urge serem investigados.

Araujo (2005), para citarmos um dos estudos que trabalharam com o tema, ao estudar a expressão de posse no português afro, encontrou construções como:

- a. (...) se papai **de nós**, governo, num dé esse dinhêro, nós num come.

A existência de tais expressões podem ser interpretadas como indícios de um passado marcado pelo contato entre línguas, uma vez que crioulos como os da ilha São Vicente e de Sotavento, em Cabo Verde, apresentam usos semelhantes. Vestígios como este, porém, podem ter sido mitigados em decorrência dos processos de urbanização e de escolarização que incidiram tardiamente sobre a população mais desvalida, ressaltando ainda mais a importância dos estudos de comunidades afro-brasileiras remanescentes.

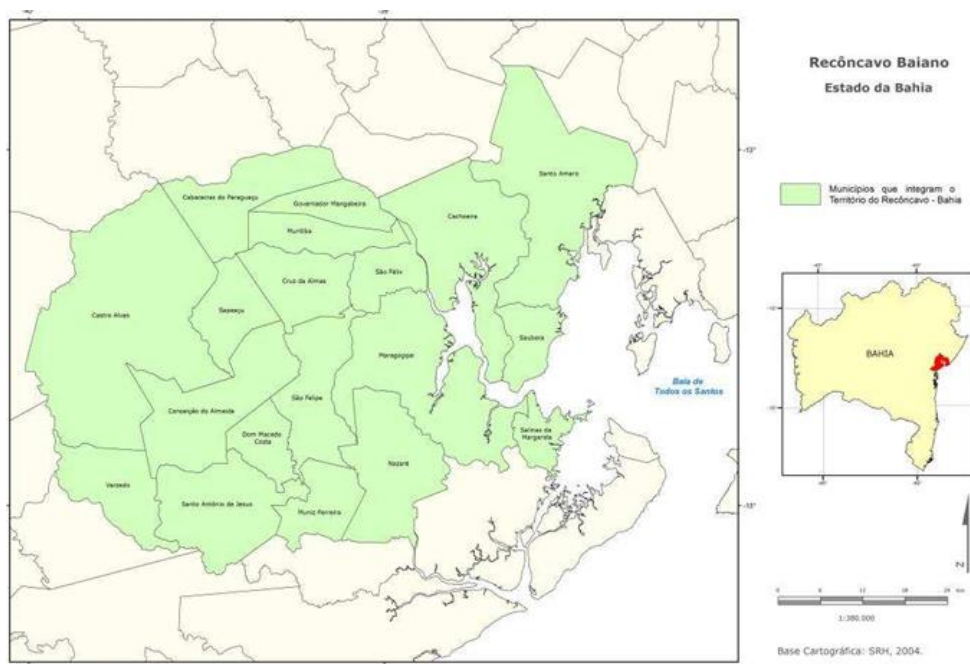
2.1 COMUNIDADES QUILOMBOLAS NO RECÔNCAVO BAIANO

Consideramos as comunidades rurais afro-brasileiras como importantes fontes de pesquisa, pois podem revelar tanto a realidade sociolinguística atual do Brasil quanto fornecer indícios de sua formação histórica. Este estudo está centrado na participação africana na formação da realidade sociolinguística brasileira, sendo o contato linguístico o fator fundamental para tal constituição. Holanda (1963) ressalta como motivos principais para a relevância de pesquisas nessas comunidades o fato de o Brasil ter sido, durante muito tempo, um país caracterizado pelas práticas rurais, e pelo fato de que essas comunidades reúnem uma expressiva concentração de pessoas de etnia não-branca e com ausência de escolarização.

Embora as comunidades remanescentes de quilombos na África tenham se constituído como espaço de resistência às políticas de escravização, com feições de iniciação militar para derrotar grandes reinos, no Brasil o cenário se mostra diverso. Em nosso país, o termo evidencia espaços transculturais, “de migrações em busca de novos territórios e de alianças políticas entre grupos alheios” (Munanga, 1995, p. 58). Araujo e Araújo (2023, p. 218) ressaltam que no Brasil, o quilombo, ao perder suas feições militares, abrigava não apenas negros fugidos do trabalho forçado, mas também indígenas, configurando-se “como um espaço transétnico e plural que abrigava diversos fugitivos que resistiam e buscavam integrar-se à resistência” (p. 8).

A Superintendência de Estudos Econômicos e Sociais da Bahia (SEI) elaborou o mapa do território de identidade do Recôncavo da Bahia (2013). Atualmente a região é composta por 19 municípios, sendo eles: Cabaceiras do Paraguaçu, Cachoeira, Castro Alves, Conceição do Almeida, Cruz das Almas, Dom Macedo Costa, Governador Mangabeira, Maragogipe, Muniz Ferreira, Muritiba, Nazaré, Santo Amaro, Santo Antônio de Jesus, São Felipe, São Félix, Salinas da Margarida, Sapeaçu, Saubara e Varzedo, conforme a figura 2, abaixo:

Figura 2 – Municípios que compõem o Território de Identidade do Recôncavo Baiano (2015)



Fonte: Elaboração NEDET (2017)

Anjos (1999) realizou um levantamento das comunidades remanescentes quilombolas registradas no Brasil. Na região Nordeste, o autor computou um total de 511, distribuídas da seguinte forma: 250 na Bahia, 163 no Maranhão, 23 em Sergipe, 16 em Pernambuco, 16 em Alagoas, 14 no Rio Grande do Norte, 11 na Paraíba, 11 no Ceará e 07 no Piauí (Anjos, 1999, 2004). Anos mais tarde, o levantamento da Fundação Cultural Palmares⁶ (FCP) (2004-2021) apresentou o seguinte número de comunidades na região Nordeste: 829 na Bahia, 845 no Maranhão, 37 em Sergipe, 196 em Pernambuco, 71 em Alagoas, 33 no Rio Grande do Norte, 47 na Paraíba, 55 no Ceará e 94 no Piauí, conforme tabela abaixo.

⁶ O aumento expressivo do número de comunidades certificadas como remanescente de quilombos deu-se em virtude do Decreto Presidencial 3.912/2001, que fixava a limitação de reconhecimento das propriedades de terras que eram ocupadas por quilombos, e pelo Decreto 4.887/2003, que regulamentou os procedimentos para identificação, reconhecimento, delimitação, demarcação e titulação das terras ocupadas por remanescentes das comunidades de quilombos, conforme disposto no art. 68/ADCT e nos artigos 215 e 216 da Constituição Federal. Passa-se então à Fundação Cultural Palmares (FCP) a competência pela emissão de certidão às comunidades quilombolas e sua inscrição em cadastro geral.

Tabela 1 – Distribuição das Comunidades Remanescentes de Quilombos (CRQs) por região

UF	Nº CRQs (Certidões)	Nº CRQs (Comunidades)
Nordeste	1.736	2.206
Sudeste	461	558
Norte	300	369
Sul	191	191
Centro-oeste	151	169
Total	2.839	3.495

Fonte: Fundação Cultural Palmares (2021)

Araujo e Araújo (2023) identificaram, a partir do levantamento realizado por Anjos (1999), o relatório da BAHIA/ SEDUR (2012), o levantamento do MDS (2015) e a tabela da FCP (2021), 73 comunidades rurais afro-brasileiras remanescentes de quilombos, localizadas na região do Recôncavo Baiano, de acordo com o quadro 1.

Quadro 1- Comunidades rurais afro-brasileiras remanescentes de quilombos Recôncavo Baiano

Municípios	Comunidades
Cachoeira	Acutinga, Brejo do Engenho da Guaiba, Caimbongo, Caiole, Calemba, Calolê, Campinas, Caonge, Coimbofo, Dendê, Desterro, Engenho da Cruz, Engenho da Pedra, Engenho da Ponte, Engenho da Praia, Engenho da Vitória, Engenho Novo do Vale do Iguape, Guaiba, Imbiara, Kaimbongo, Kalembe, Kaonje, Mutecho, Opalma, Santiago do Iguape, São Francisco do Paraguaçu, São Tiago do Iguape, Tabuleiro da Vitória e Tombo.
Cruz das Almas	Baixa da Linha e Vila Guaxinin
Maragogipe	Angolá, Baixão do Guai, Buri, Enseada do Paraguaçu, Giral Grande, Guarucú, Guerém, Pinho, Porto da Pedra, Quizanga, Salamina Putumuju, Salaminas, Sítio Dende, Tabatinga, Topa de Cima e Zumbi.
Muritiba	Baixa Grande
Santo Amaro	Alto do Cruzeiro-Acupe, Barro Velho, Barro Vermelho, Caeira, Cambuta, São Braz e Subaé
Santo Antônio de Jesus	Alto do Morro, São Roque dos Macacos e Terra Seca
São Félix	Engenho São João, Engenhoca, Oiteiro Redondo, Pau Grande, Quilombo, Santo Antônio, Santo Antônio da Jaqueira, Subahuma, Terr[enos] do Governo e Vidal

São Francisco do Conde ⁷	Ilha do Paty, Monte Recôncavo e Porto de Dom João
São Sebastião do Passé	Fazenda Nossa Senhora do Espírito Santo, Palmeira da Água Boa

Fonte: Araujo e Araújo (2023), adaptado de Anjos (1999), BAHIA/ SEDUR (2012), MDS (2015) e FCP (2021).

Diante da grande concentração de comunidades quilombolas na Bahia, torna-se, portanto, relevante o desenvolvimento de pesquisas nessas comunidades para o conhecimento do PB a partir de pesquisas empíricas, de modo a revelar possíveis efeitos ainda preservados do contato entre a língua portuguesa e as centenas de línguas africanas transplantadas para o Brasil. Dentre as comunidades no Recôncavo Baiano que receberam a certificação de remanescente quilombola, duas localizam-se na cidade de Cruz das Almas: a comunidade Vila Guaxinin e a comunidade Baixa da Linha. No item seguinte apresentamos algumas informações referentes à comunidade analisada nesta pesquisa, a Baixa da Linha.

2.1.1 A Comunidade quilombola Baixa da Linha

A comunidade quilombola Baixa da Linha está localizada no município de Cruz das Almas-BA, no Recôncavo Baiano. O município foi uma freguesia de Cachoeira, cidade importante economicamente no período Colonial e no Império, por ser o último ponto navegável do Rio Paraguassu e entrada para as rotas de tropeiros em direção ao sertão, além de ser produtora de cana-de-açúcar e fumo. A Freguesia de Nossa Senhora do Bom Sucesso da Cruz das Almas, datada de 1815, origina-se em um ponto da rota de tropeiros com destino as terras do sertão, onde havia uma cruz em que pessoas se reuniam para “fazer novenas, pagar promessas, invocar os santos e rezar pelas almas”, surgindo desta forma, sua denominação (Cunha, 1959, p. 28). A freguesia de Cruz das Almas, apesar de não produzir cana-de-açúcar tinha a importante missão de produzir lavoura de subsistência e fumo, o que atraiu os primeiros povoadores do território, as famílias Rocha Passos e Batista Magalhães, ambas descendentes de portugueses que mantinham em suas fazendas africanos escravizados (Cunha, 1959). Com possíveis finalidades políticas, um membro da família Rocha Passos divulga uma notícia no jornal *Gazeta da Bahia*, em 1883:

⁷ Em 2015, os municípios de São Francisco do Conde e São Sebastião do Passé migraram para o Território da Região Metropolitana de Salvador, e o município de Salinas da Margarida migrou para o Território de Identidade (TI) do Recôncavo Baiano. Neste município há uma comunidade certificada pela FCP, a comunidade pesqueira Conceição de Salinas.

Carta de liberdade – Eu, abaixo assignado, desejando concorrer para a emancipação gradual e pacífica dos captivos, e não me sendo possível, em consequência de minhas circunstancias e onerados da família, libertar de uma vez um numero crescido dos que possuo, tenho, todavia, resolvido pelo presente documento, por mim feito e assignado, conceder carta de liberdade sem ônus ou condição alguma aos meus escravos Jacob, africano, Ascanio, africano, Brigida, africana, e Martinha, crioula, todos matriculados na coletoria geral da cidade da Cachoeira. Confiro também liberdade a escrava Romana, crioula, pela quantia de 90\$000, muito inferior ao seu valor, pois que, não podendo obter mais, entendi não privar-a deste benefício. Poderão os escravos acima gozar de sua liberdade como de ventre livre nascessem.

Engenho Campo Limpo da Freguesia de Cruz das Almas, 10 de abril de 1883. – Temistocles da Rocha Passos (Gazeta da Bahia, 1883, p. 2).

Fica evidente a presença de africanos escravizados e descendentes de europeus no território da atual cidade de Cruz das Almas. Apesar de não possuir informações precisas sobre o número de escravizados que foram forçados a habitar nesta região, há indícios de que a quantidade foi expressiva, pois o próprio autor da notícia revela possuir muitos cativos.

Os relatos dos moradores mais antigos que residem atualmente na comunidade Baixa da Linha, baseados na memória de seus ancestrais, afirmam a primeira ocupação de povos indígenas naquela área; vestígios históricos encontrados pelos moradores comprovam as narrativas dos anciãos. A figura 3 exibe fragmentos cerâmicos e vários líticos lascados, assim como urnas funerárias, com data estimada anterior a 1500.

Figura 3 – Fragmentos de cerâmica, líticos lascados e urnas funerárias



Fonte: Sítio arqueológico (UFRB)

Tem sido muito comum nos últimos anos a descoberta de urnas funerárias por moradores da comunidade, provavelmente da comunidade indígena Aratu⁸ (Silva, 2020, p. 57), cuja extinção remonta um período anterior à colonização portuguesa. A partir dessas e de outras evidências arqueológicas encontradas, estima-se que os primeiros habitantes da cidade de Cruz das Almas foram os povos indígenas, Cariris e Sabujás. Fernandes (2011) argumenta que estes grupos indígenas produtores da cultura material – cerâmica – não foram contatados⁹ pelo elemento colonizador, tendo se extinguido ou se transformado muito antes.

A CRQ que integra esta pesquisa, situada na zona rural da cidade de Cruz das Almas, foi escolhida pela sua expressiva representatividade negra resultante da descendência dos moradores atuais de pessoas submetidas à escravidão as quais encontraram refúgio da tortura nesta região. Esta é a principal narrativa transmitida pelos moradores mais antigos, o que promove uma reflexão sobre a precária situação econômica ainda vivenciada pelos moradores desta CRQ. Diante da luta por sobrevivência da população afro-brasileira e da inexistência de políticas públicas que fornecessem auxílio a esses “novos cidadãos”, os quilombos representavam o abrigo e proteção de que necessitavam. De acordo com Od’wyer (2002), desde a Constituição Federal de 1988, o termo remanescente de quilombo já vinha sendo reformulado e passou a representar uma herança cultural e material que atribui a esses grupos um sentimento de pertencimento a um território específico. Segundo a autora, esse grupo se define pelas experiências vividas, “práticas de resistência na manutenção e reprodução de seus modos de vida característicos e na consolidação de um território próprio” (p. 18). Desse modo, o Estado precisa garantir as condições necessárias para que esses grupos se desenvolvam social, econômica, física e culturalmente (Brasil, 2003). Coadunando com o pensamento de Od’wyer (2002), Moura (1996), estabelece:

(...) quilombos contemporâneos no Brasil, comunidades integradas à humanidade, cuja diversidade cultural precisa ser respeitada e reconhecida. Para o imaginário brasileiro, quilombo foram agrupamentos de africanos escravizados fugidos de engenho, fazendas e minas que tentaram reproduzir vida comunitária à semelhança da África, terra de origem, para fugir dos maus-tratos infligidos pelo senhorio branco europeu. Cabe tratar de quilombo contemporâneo comunidades remanescentes, terras de preto, mocambos,

⁸ Martin (1999, p. 186-187) definiu algumas características da cultura Aratu, como: a) utilização da cerâmica roletada, sem decoração, com as superfícies alisadas ou engobo de grafite; b) urnas funerárias piriformes, com e sem tampa, de 70-75 cm de altura; tigelas menores empregadas como opérculo para cobrir os vasilhames funerários; c) painéis semiesféricos de bordas onduladas; d) enterramentos primários em urna, fora das aldeias; dentre outras características.

⁹ Perota (1971, p. 8) registra um único vestígio europeu junto ao material arqueológico de Aratu, um pequeno caco de vidro. No entanto, não há elementos suficientes para realizar qualquer confirmação.

terras de santo ou santíssimo ou terras de herança sem formal de partilha que o movimento social fez sinônimas, aplicando o artigo 68 da Constituição. Desvela parte da história do negro no Brasil atual (Moura, 1996, p. 327-328).

Sendo assim, para a autora, os quilombos contemporâneos constituem espaços que revelam a história do negro no Brasil, uma história que precisa ser desvelada e reconhecida. Portanto, é necessário que haja um deslocamento atento de pesquisas para esses espaços, principalmente quando se busca investigar fenômenos linguísticos sensíveis ao contexto de aquisição, já que a participação da população negra foi profunda e maciça nas mais diversas esferas das atividades de implantação do português no Brasil.

A comunidade negra rural da Linha, que se localiza a quatro quilômetros do atual centro da cidade de Cruz das Almas-BA, localizada a 146 Km de Salvador, é composta atualmente por 120 famílias cujos antepassados se instalaram nas proximidades da linha do trem e porventura foram construindo suas casas. Segundo Silva (2020), a Estação Ferroviária Central da Bahia foi inaugurada em 23 de dezembro de 1881, anos mais tarde passou a ser chamada de Estação Ferroviária de Cruz das Almas, atraindo na década de 1930 uma pedreira que se instalou nos arredores da comunidade. A instalação da Estação Ferroviária deu-se nove anos antes da Abolição da Escravidão no Brasil, o que pode explicar a vinda de africanos fugidos e seus descendentes para a comunidade da Linha. Nesse local, os grupos sobreviveram mediante atividades agrícolas para o abastecimento das suas famílias. Inicialmente, havia apenas casas de taipa e de palhas, e foi desse modo que a comunidade se desenvolveu no espaço próximo à linha férrea.

O trem não era apenas um dos meios privilegiados de transporte no município de Cruz das Almas, mas também uma rota de fuga, de interlocução, de troca cultural e de formas humanas de sobrevivência, adentrando as fazendas e matas. A proposta tipológica apresentada por Santana, Araujo e Freitag (2018) para as CRQ em Sergipe, envolve parâmetros sociolinguísticos como os contatos, deslocamentos populacionais e a relação entre português rural e o PAF. Esses autores, estendem o critério de representatividade de Lucchesi, Baxter e Ribeiro (2009) de comunidades [+ isoladas] e [+ afrodescendentes] para [$\pm$ isoladas], [+ afrodescendentes], [+ africanizadas] e localizadas em municípios (microáreas) e regiões (macroáreas) que foram, nos períodos colonial e imperial, [+ escravocratas].

Seguindo esta tipologia, precisamos descrever a representatividade da comunidade da Linha. Esta comunidade não pode ser considerada totalmente isolada, pois, está situada na área posterior à Universidade Federal do Recôncavo da Bahia (UFRB), conforme figura 4, e possui

como característica o fluxo de alguns moradores, principalmente adultos, do centro da cidade para a comunidade, motivado por questões laborais. Este fato corrobora com a afirmação de Santana, Araujo e Freitag (2018) sobre as comunidades remanescentes de quilombos, pois “devido a mudanças nas suas dinâmicas sociais, não estão totalmente isoladas” (Santana, Araujo e Freitag, 2018, p. 208). Os membros da comunidade estão expostos à mídia, devido ao auxílio de voluntários para que eles tivessem acesso não só à energia elétrica como também à água encanada. Prosseguindo à classificação, a comunidade pode ser representada como:

- a. [+afrodescendente], pois apesar da escassez de fontes históricas, as narrativas dos informantes mais antigos revelam que seus ancestrais eram descendentes de africanos escravizados que resistiram ao sistema escravocrata ao qual foram submetidos em grandes fazendas e engenhos nas cidades circunvizinhas e históricas de Cachoeira, São Félix e Santo Amaro da Purificação para se abrigarem em pontos estratégicos, como nas proximidades da linha do trem.
- b. [+ africanizada] no sentido em que parte dos membros estão vinculados às práticas culturais da comunidade, como as práticas religiosas de terreiros que acontecem na própria CRQ e a pontual participação no samba de roda, conforme figura 5. No entanto, outra parcela da comunidade desconhece os efeitos da escravidão de seus antepassados que reflete diretamente na atual conjuntura em que vivem.
- c. [+ escravocrata] pois a região do Recôncavo Baiano foi um grande pilar para o desenvolvimento da produção açucareira, chegando a possuir no final do século XVIII mais de 215 engenhos.

Assim quanto maior o grau de atuação desses parâmetros, maior será a TLI na formação da variedade linguística falada por seus moradores (Santana, Araujo e Freitag, 2018, p. 125-126). Assim, os dados desta pesquisa contribuem para confirmar a relevância dos parâmetros sociolinguísticos propostos por Santana, Araujo e Freitag (2018), demonstrando que a configuração linguística das comunidades quilombolas está diretamente relacionada às condições históricas de formação, aos níveis de isolamento social e às dinâmicas de contato linguístico que caracterizam essas comunidades.

Figura 4 – Vista área da Universidade Federal do Recôncavo Baiano (UFRB)



Fonte: TV UFRB (2021)

Desse modo, o passado de escravização, bem como suas implicações sociais e linguísticas, se mantém vivo nessa comunidade, principalmente por meio dos relatos orais dos moradores mais antigos:

Inf.: Eu conhecia aqui uma senhora que ela quando apareceu aqui, ela tinha vinte ano, chegou aqui com vinte ano... uma... duas... quatro... e quatro filha, aí ela ficou aqui, porque ela saiu corrida naquele tempo. Ela saiu corrida, lá da onde ela vivia porque tinha que trabalhar de dia e noite e no dia que ela pôde ela saiu, ela saiu corrida e apareceu aqui (SS03).

Desse modo, o informante SS, da faixa etária 3, confirma que aquele espaço abrigava pessoas submetidas ao regime da escravidão, possivelmente nos engenhos entre Cachoeira e Santo Amaro ou alguma fazenda na região do Recôncavo.

A principal atividade desenvolvida pela comunidade como meio de subsistência é a agricultura, sendo impulsionada pelas pesquisas no âmbito da Universidade Federal do Recôncavo da Bahia (UFRB), a qual também tem fomentado projetos de incentivo ao desenvolvimento de uma identidade quilombola na comunidade. Os estudos de Cunha (1995, 2003) identificam traços linguísticos associados ao contato entre o português e línguas africanas, especialmente em comunidades negras rurais isoladas, e demonstram que as comunidades quilombolas são espaços fundamentais para compreender a diversidade linguística do português brasileiro, as heranças africanas na língua e na cultura, bem como a relação entre língua, memória e identidade afro-brasileira. Segundo a autora, as comunidades

quilombolas constituem espaços privilegiados para observar a interação entre língua, memória e identidade afro-brasileira, evidenciando traços de variação linguística e possíveis vestígios do contato entre o português e línguas africanas no português rural dessas comunidades.

Em comunidades quilombolas, práticas culturais como o samba assumem papel central na construção da identidade coletiva e na preservação de heranças culturais africanas, articulando memória, tradição e pertencimento social (Cunha, 2003), conforme figura 5:

Figura 5- Apresentação cultural de samba de roda na comunidade Baixa da Linha



Fonte: Elaborado pela autora (2024)

Nesse contexto, o samba não deve ser compreendido apenas como manifestação artística ou forma de entretenimento na comunidade, mas como uma prática social que desempenha funções simbólicas e históricas fundamentais. Por meio da música, da dança e da oralidade que caracterizam essa manifestação cultural, são transmitidos conhecimentos, valores e experiências coletivas que reforçam a memória histórica dessas comunidades. Além disso, o samba constitui um espaço de resistência cultural, no qual se reafirmam vínculos comunitários e identidades afro-brasileiras construídas ao longo de processos históricos marcados pela escravidão, pela marginalização social e pela luta por reconhecimento. Dessa forma, tais práticas culturais contribuem para a manutenção e a atualização de tradições herdadas de matrizes africanas, ao mesmo tempo em que fortalecem o sentimento de pertencimento e a continuidade cultural nas comunidades quilombolas.

A antiga estação, próxima à comunidade, transportou muitas pessoas para as várias regiões do Brasil e também realizava o processo de escoamento de matéria prima ali produzida, sendo fundamental para a economia da cidade. Atualmente, a estação encontra-se abandonada, conforme figura 6, mas foi no entorno dessa via, que se desenvolveu um grupo cujos laços ainda persistem.

Figura 6 – Antiga Estação de trem de Cruz das Almas



Fonte: Nilton Antonio Souza (2009)

O transporte ferroviário, nessa perspectiva, deve ser entendido tanto pelo ponto de vista da ascensão da economia brasileira no final do século XIX, com o transporte de mercadorias como farinha, café, açúcar, fumo, dentre outros, quanto pelo transporte de pessoas, incluindo africanos escravizados fugidos das fazendas, engenhos e minas que buscavam reproduzir uma vida comunitária à semelhança da África. A figura 7 exhibe um dos cenários da comunidade.

Figura 7 - Vista da Comunidade Baixa da Linha



Foto: Lima (2015)

O conceito remanescente de Quilombo se relaciona à existência ou sobrevivência na comunidade de elementos do seu processo histórico do passado, uma característica que encontramos na comunidade Baixa da Linha. Assim, para Reis (1996):

No Brasil essa memória não é absolutamente apagada, se consideradas as comunidades chamadas de remanescentes de quilombos que realmente podem traçar seu passado até agrupamentos constituídos antes da abolição, em 1888 (Reis, 1996, p. 10).

A observação de Reis (1996) chama atenção para a permanência de elementos da memória histórica da população negra no Brasil, sobretudo quando se consideram as CRQs. Embora o processo de escravização e as transformações sociais posteriores à abolição tenham produzido rupturas profundas nas formas de organização social e cultural dessas populações, essa memória coletiva não foi completamente apagada. Em muitas comunidades quilombolas, é possível identificar vínculos históricos que remontam a agrupamentos formados ainda no período escravista, o que evidencia a continuidade de práticas sociais, culturais e territoriais ao longo do tempo. Nesse sentido, tais comunidades constituem espaços privilegiados para a preservação de experiências históricas e identitárias da população afro-brasileira. Além de sua relevância histórica e social, esses contextos também se mostram particularmente importantes para os estudos linguísticos, uma vez que podem conservar traços de variedades do português que se desenvolveram em situações específicas de contato linguístico. Dessa forma, a permanência dessa memória histórica contribui para compreender não apenas a formação dessas comunidades, mas também os processos socioculturais e linguísticos que marcaram a

constituição do PB em contextos afrodescendentes. A figura 8 exibe um espaço importante para a comunidade estudada.

Figura 8 – Associação comunitária quilombola da Baixa da Linha



Fonte: Roger Conrado (2018)

A Associação Comunitária foi criada em 1998 e sua sede foi construída no ano 2000, conforme figura 8, pela Igreja Católica com o apoio de comerciantes locais. A criação da Universidade Federal do Recôncavo da Bahia (UFRB), em 2005, cuja reitoria localiza-se na cidade de Cruz das Almas, gerou novos embates referentes à garantia de terras para as comunidades quilombolas que residem nesta região.

Somente em 2010 a comunidade Baixa da Linha conquista o Reconhecimento e a Certificação como comunidade quilombola pela FCP. Em 2019, com a Demarcação das Terras das comunidades e a delimitação das terras pertencentes à Universidade e aos quilombolas, se encerraram anos de conflitos e debates acerca do território de direito dos povos tradicionais. Assim, segundo Lucchesi (2009, p. 32), o PAF guarda, nessas comunidades, uma especificidade única no universo mais amplo do português popular rural brasileiro (ou, mais precisamente, norma popular rural do português brasileiro), o que reforça ainda mais a necessidade da sua preservação e da garantia de seus direitos.

3 A REALIZAÇÃO DO SUJEITO PRONOMINAL EM VARIEDADES DO PORTUGUÊS

O português, tomado em seu conjunto, traz questões particularmente instigantes para o sintaticista, porque ele apresenta fenômenos ausentes das outras línguas românicas, pelos menos as mais estudadas (Galves, 2001, p. 15).

Esta seção é dedicada a retomar estudos sobre o sujeito pronominal em algumas variedades do português, com foco em apontar os desdobramentos de algumas teorias e tecer comparações para auxiliar a reflexão sobre a realidade brasileira, especificamente sobre o PAF, no que tange ao sujeito pronominal no âmbito da TPP. Para tanto, a seção está subdividida nos seguintes itens: Pronomes pessoais – algumas discussões, na qual explicamos em que consiste o sujeito pronominal, realizando uma breve revisão sobre os pronomes bem como às pessoas pronominais no PE, no português brasileiro (PB) e no português afro-brasileiro¹⁰ (PAF). Na seção seguinte, O parâmetro do sujeito nulo, dedicada ao PSN, contextualizamos a TPP, seu surgimento e suas contribuições, além de esboçarmos algumas línguas que exibem uma marcação positiva para esse parâmetro. Por fim, nas subseções do item 3.3, O parâmetro do sujeito nulo em variedades do português, é demonstrado como esse parâmetro se manifesta nas variedades europeia, brasileira e afro-brasileira do português.

3.1 Pronomes Pessoais: Algumas Discussões

Antes de discutirmos o sujeito pronominal, convém descrever como os pronomes são identificados em variedades do português. As gramáticas descritivas do latim já identificavam os pronomes como integrantes da categoria gramatical atuantes nas situações de discurso. Esse legado latino se faz fortemente presente na tradição gramatical do PB, cujo quadro pronominal é derivado da flexão de casos do latim. Assim, os pronomes retos (caso “nominativo” do latim)

¹⁰ O conceito de português afro-brasileiro para a realidade brasileira fundamenta-se nos estudos de Lucchesi (2009, p. 31-32), o qual afirma que o termo “designa uma variedade constituída pelos padrões de comportamento linguístico de comunidades rurais compostas em sua maioria por descendentes diretos de escravos africanos que se fixaram em localidades remotas do interior do país, praticando até os dias de hoje a agricultura de subsistência”. O conceito de português afro-brasileiro se fundamenta não em parâmetros étnicos, mas em parâmetros sócio-históricos.

ocupam a função de sujeito, e os pronomes oblíquos (casos “acusativo” e “dativo”) têm função de complemento.

Tradicionalmente, o pronome é definido como substituto do nome, que fica em lugar ou função de um nome. Monteiro (1994) argumenta contra a imprecisão do conceito de pronome como substituto dos nomes, pois há outras classes gramaticais que podem fazê-lo. Lopes (2013, p. 107) confirma a ideia de Monteiro (1994) sobre a impossibilidade de definir a classe como “substituição dos nomes”, acrescentando que esta propriedade não se aplica a toda classe de pronomes, sendo restrita a alguns deles¹¹. Bhat (2004) também problematiza a definição tradicional dos pronomes e propõe que eles sejam um rótulo sob o qual se incluiriam os pessoais prototípicos (primeira e segunda pessoa) e proformas (demais pronomes). Bhat (2004) inclui também outras propriedades desta classe além da condição de participante do discurso (falante/ouvinte), como o papel temático e caso.

Apesar desses esforços, esses critérios não são universais, dado o comportamento singular desta categoria. Chomsky (1981), Galves (2001), dentre outros, argumentam sobre a arbitrariedade dos pronomes pessoais, uma vez que estes podem apresentar um comportamento semântico semelhante ao de um nome genérico. Nos exemplos (1a-b), extraído do inquérito 13 da comunidade afro-brasileira de Helvécia, que faz parte do banco de dados do Projeto Vertentes do Português Popular do Estado da Bahia (Lucchesi, 2002), os pronomes *eu*, *você* e *a gente* indicam que neste contexto sintático toda e qualquer pessoa que começa a fumar cigarro hoje irá experimentar “X” amanhã. Sendo assim, esses pronomes de primeira e segunda pessoa com leitura genérica são pronomes arbitrários, pois não identificam alguém particular que começa a fumar.

- (1) a. *Eu* começo a fumar um cigarro hoje, amanhã *eu* experimento X.
 b. *Você/a gente* começa a fumar um cigarro hoje, amanhã *você/a gente* experimenta X.

Desse modo, um pronome arbitrário assume uma leitura genérica, pois não é possível identificar sua referencialidade, ou seja, os pronomes em (1a-b) não possuem nem o traço de definitude, nem o de especificidade. Mateus *et al* (2003) afirma que os pronomes pessoais

¹¹ Lopes (2013, p. 107) argumenta que formas pronominais que se caracterizam como determinantes, como os possessivos, não podem substituir um nome: (*Maria/ela/aquela/alguém/*minha* foi à festa). A autora salienta outro aspecto dessa “substituição”, uma vez que ela não é necessariamente do nome, mas de todo o SN. Quando o sintagma tem como núcleo apenas um nome, a substituição será exclusiva: (*Patrícia* viajou x *ela/alguém* viajou). Em outro caso, o pronome tomará o lugar de todo o sintagma: (*Aquela esperta menina de três anos* viajou sozinha x *ela/ alguém* viajou sozinha).

“denotam a pessoa gramatical das entidades participantes no ato de comunicação (locutor, ouvinte e entidade acerca da qual se fala)” (p. 826). Outro estudo que merece considerável atenção é o desenvolvido por Émile Benveniste (1956) que desenvolve uma profícua discussão sobre o agrupamento das três pessoas do discurso. Para ele, pessoas do discurso refere-se apenas àqueles que participam desse processo de forma efetiva. Assim, a noção de pessoa se aplica apenas às primeiras e segundas pessoas que são intercambiáveis. Para o referido autor, a terceira pessoa não possui tal particularidade, pois esta se encontra fora da situação comunicativa, e além disso, a terceira pessoa não necessariamente faz referência a um ser humano, como a primeira e a segunda pessoa.

Monteiro (1994) coaduna com as ideias de Benveniste (1956), mas ressalta, sobre a categorização ideal das pessoas do discurso, ser “inviável [...] propor uma nova designação ou tentar uma sistematização mais adequada” (Monteiro, 1994, p. 32), tendo em vista a incoerência de vocábulos como *eu* e *ele* quando estes possuem funções e traços morfossintáticos diversos. Sob o aspecto da referência dos pronomes pessoais, Bloomfield (1938), Forchheimer (1953), Benveniste (1966) e Bhat (2004) contestaram até que ponto o pronome de terceira pessoa pode ser considerado de fato um pronome pessoal, uma vez que a terceira pessoa não é estabelecida na enunciação com locutor e interlocutor assumindo esses papéis e alternando-se no decorrer do discurso.

Sem pensar em esgotar esta discussão com respeito às propriedades dos pronomes, convém refletir a perspectiva defendida por Castilho (2010, p. 475), para ele, as línguas naturais priorizaram “a codificação (i) das pessoas do discurso; (ii) dos lugares ocupados por elas no espaço físico; e (iii) de seu tempo”. Essas propriedades assegurariam o caráter dêitico dos pronomes. No entanto, o pronome pessoal de terceira pessoa faz surgir uma outra propriedade, a de retomada de elementos no texto, também conhecida como anáfora. O autor conclui que muito longe de acarretar uma perda de propriedades aos pronomes, a terceira pessoa contribuiu para um acréscimo de propriedades.

Buscamos destacar, nesta subseção, o comportamento peculiar do pronome de terceira pessoa, pois, conforme pesquisas citadas anteriormente, ele constitui, em relação à pessoa do discurso, o maior contexto de resistência à expressão do sujeito pleno no PB, haja vista que a sua identificação se dá pela correferência com algum sintagma nominal, pelo contexto discursivo ou pragmático (Duarte, 1995, 2019). Conforme os dados de Duarte (1995), os pronomes no PB vêm demonstrando, na função de sujeito (caso nominativo), uma crescente realização fonética, a partir do final do século XIX. A seguir, exibiremos alguns estudos que

destacaram as pessoas pronominais nominativas no português europeu, brasileiro e afro-brasileiro.

3.1.1 Quadro pronominal em variedades do português: europeu, brasileiro e afro-brasileiro

No latim clássico, havia duas formas específicas para os pronomes nominativos de primeira e segunda pessoas. Almeida (2000) aponta que não havia uma forma específica para o pronome pessoal de terceira pessoa, ou seja, não havia um pronome pessoal que fizesse referência externa na enunciação. O surgimento do pronome *ele*, portanto, tem sua origem na transição do latim para o português, com a incorporação do demonstrativo *ille* entre os pronomes pessoais do caso nominativo (Loregian-Penkal, 2004, p. 29-30).

A proposta de análise de tais pronomes realizada por Mateus *et al* (2003), conforme tabela 2 abaixo é de suma importância para o entendimento da terceira pessoa do discurso. A referida autora apresenta os pronomes no PE e classifica-os quanto ao seu valor dêitico ou correferencial.

Tabela 2: Pronomes nominativos no português europeu (PE)

Pessoas gramaticais	Apenas valor dêitico	Valor dêitico/correferencial
1ª pessoa do singular	Eu	--
2ª pessoa do singular	Tu/você	--
3ª pessoa do singular	--	Ele/ela
1ª pessoa do plural	Nós	--
2ª pessoa do plural	Vós/vocês	--
3ª pessoa do plural	--	Eles/elas

Fonte: Mateus *et al*, 2003, p. 819.

Para a autora, no PE, o valor dêitico ou anafórico apenas pode ser representado pelos pronomes *eu*, *tu/você*, *nós*, *vós/vocês*, pois esses pronomes só podem assumir um valor referencial fixado no sintagma nominal. Já os pronomes de terceira pessoa podem assumir dois valores, referenciando tanto um elemento presente no sintagma, em que estabelece correferencialidade, quanto uma entidade fora da situação comunicativa. A identificação desse referente envolve, portanto, fatores semânticos e textuais/discursivos, para além do domínio sintático.

Para Mateus *et al* (2003), a correferência está relacionada à capacidade de um pronome ou categoria vazia se relacionar à expressão nominal que o antecede. Desse modo, é possível perceber que o pronome de terceira pessoa possui peculiaridades que o distinguem dos demais pronomes pessoais, uma vez que possui dois valores, diferentemente dos pronomes de primeira e segunda pessoa. A abordagem de Castilho (2010), de acordo com a tabela 3, para o quadro pronominal do PB, nos permite perceber uma diversidade de formas a depender de fatores sociais e contextuais.

Tabela 3: Quadro pronominal do português brasileiro (PB)

Pessoas gramaticais	FORMAL	INFORMAL
1ª pessoa do singular	Eu	Eu, a gente
2ª pessoa do singular	Tu, você, o senhor, a senhora	Você, ocê, cê, tu
3ª pessoa do singular	Ele, ela	Ele/ei, ela
1ª pessoa do plural	Nós, a gente	A gente
2ª pessoa do plural	Vós, os senhores, as senhoras	Vocês, ocês, cês
3ª pessoa do plural	Eles, elas	Eles/eis, elas

Fonte: Castilho, 2010, p. 477

Com respeito à primeira pessoa do plural é possível observar que o uso de *a gente* não se limita a contextos informais, mas vem expandindo sua frequência de uso concorrendo com o pronome “nós” em contextos formais. Sobre o processo de gramaticalização de “a gente”, Lopes (1999, p. 1) explica que o substantivo “gente” (sinônimo de “as pessoas”) foi cristalizado na forma “a gente”, integrando o sistema pronominal como indicação da primeira pessoa plural em variação com “nós”. O percurso percorrido por esta forma que, segundo Lopes (1999), se origina por volta do século XIII de forma generalizante e indeterminada, vai de *gente* [nome genérico] > *a gente* [pronome indefinido] > *a gente* [pronome pessoal].

Já em relação à segunda pessoa, vem ocorrendo uma substituição do pronome *tu* pelo *você*, cuja forma é resultante da gramaticalização da forma de tratamento *vossa mercê*. Dentre as regiões que preservam a forma conservadora *tu*, a utilização de *você* denota distanciamento, principalmente no PE. Castilho (2010) também destaca o uso de *o senhor/ a senhora* como pronomes pessoais para tratamento cerimonioso. O caso que mais chama a atenção, apontado pelo autor, são as formas *ei, eis* utilizadas como variantes, em contextos informais, referindo-

se, respectivamente a *ele/eles*. A hipótese do autor para tais usos é de ordem fonética, uma vez que *ele/eles* são dissilábicos, a redução ocorra como uma tentativa de aproximação com os demais pronomes monossilábicos *eu*, *tu* e *cê*.

Comparando o quadro pronominal do PB apresentado por Castilho (2010) com o do PAF¹² levantado por Lucchesi (2009), é possível encontrar muitas semelhanças.

Tabela 4: Quadro pronominal do português afro-brasileiro (PAF)

Pessoas gramaticais	Sujeito
1ª pessoa do singular	Eu
2ª pessoa do singular	Tu, você, ocê, cê
3ª pessoa do singular	Ele, ela
1ª pessoa do plural	Nós, a gente
2ª pessoa do plural	Vocês
3ª pessoa do plural	Eles, elas

Fonte: Lucchesi, 2009, p. 481.

A redução fonológica constitui uma característica recorrente em línguas que se formaram a partir do contato linguístico, nos quais a intensidade e a duração do contato podem provocar mudanças significativas na estrutura gramatical. No caso do PB, tais processos não se restringem à segunda pessoa do singular, mas afetam diferentes pessoas do discurso e diversas formas verbais.

Conforme observa Castilho (1998, 2010, 2019), a história e o funcionamento do PB evidenciam processos de simplificação fonológica e morfológica que atingem o sistema pronominal e a flexão verbal como um todo. Um exemplo conhecido é a evolução de *vossa mercê* para *você*, frequentemente reduzida na fala para *ocê* ou *cê*, fenômeno que contribuiu para a reorganização do paradigma pronominal. No entanto, reduções semelhantes podem ser observadas também em outras pessoas do discurso, como na primeira pessoa do plural, em que formas como *estamos* e *vamos* são frequentemente realizadas como *tamo(s)* e *vamo(s)* em variedades populares, bem como em estruturas em que a marca de plural verbal é enfraquecida ou reduzida, como em *nós fala* ou *a gente fala*. Processos de redução também atingem formas verbais de terceira pessoa, como em *está - tá* e *estava - tava*, além da redução de preposições e conectores, como *para - pra*. Esses fenômenos ilustram tendências de simplificação fonológica

¹² Consideramos português afro-brasileiro nesta pesquisa, a variedade de língua portuguesa falada por moradores de comunidades descendentes de escravizados que permaneceram isolados no interior do país até meados do século XX (Lucchesi, 2009, p. 75).

e morfológica que, segundo Castilho, fazem parte dos processos naturais de mudança linguística, mas que podem ser intensificados em contextos históricos marcados pelo contato entre línguas, como ocorreu na formação do português brasileiro. Dessa forma, a redução fonológica deve ser compreendida como um processo mais amplo, que afeta diferentes componentes da gramática e contribui para a reorganização do sistema pronominal e da morfologia verbal do PB.

O quadro pronominal da variedade afro-brasileira do português, conforme tabela 4, é muito semelhante com o português informal apresentado por Castilho (2010), o que pode ser explicado pelos “processos de mudança que se difundem numa mesma direção por todos os segmentos da sociedade brasileira, como a realização do sujeito pronominal, o objeto nulo, as relativas cortadoras etc.” (Lucchesi, 2015, p. 267).

O sujeito, nesse sentido, é uma categoria gramatical que suscita muitos questionamentos. Dentre gramáticos que procuram analisar o seu funcionamento, destacamos as contribuições de Castilho (2010), cujo estudo descritivo, baseado em linguística de *corpus*, levanta discussões importantes para se compreender suas propriedades. O autor discute sobre as possibilidades de preenchimento do sujeito, afirmando que essa posição pode ser ocupada por diferentes classes, como substantivos, expressões pronominais ou por uma sentença, destacando que a posição de sujeito pode ser “ocupada” também por uma categoria vazia. Já a abordagem de Duarte (2007) chama a atenção para o uso da nomenclatura “essencial”, principalmente na Nomenclatura Gramatical Brasileira, que é utilizada para se referir ao sujeito, apontando uma maior importância no contexto oracional que os demais termos, pois, segundo a autora, esta denominação não traz contribuições para o entendimento do funcionamento da categoria sujeito. A autora não considera, por isso, apenas o sujeito e o predicado como termos essenciais, mas também o predador, o termo com o qual os demais termos se articulam. Assim, amparada pela teoria gerativa, o sujeito é entendido como argumento externo, cuja existência está ancorada à seleção feita pelo predador.

Como conclusão diante desse cenário, a autora defende ser de grande importância que os alunos reconheçam e identifiquem os constituintes da sentença, não só para a compreensão da concordância, mas também “para a boa utilização dos sinais de pontuação” (Duarte, 2007, p. 204). As definições e casos abordados pela gramática tradicional não dão conta, assim, de toda complexidade que envolve as línguas humanas. Para uma compreensão mais profunda sobre a categoria sujeito pronominal, é fundamental, portanto, analisá-lo sob vários aspectos, sejam sintáticos, semânticos ou discursivos.

3.2 O Parâmetro do Sujeito Nulo

Desde o início do desenvolvimento da Teoria Gerativa, a concepção de que há nos seres humanos uma capacidade inata para a aquisição da linguagem, denominada de Gramática Universal (GU), tem suscitado o desenvolvimento de diversas pesquisas. Segundo o modelo gerativista, a GU é composta por um conjunto de princípios, que são propriedades invariáveis e comuns a todas as línguas humanas, e parâmetros, que correspondem a propriedades subespecificadas, sendo assim, passíveis de variação (Kato, 2002, p. 311). Assim, as línguas possuem regularidades gramaticais universais e ao mesmo tempo, variações linguísticas possíveis (Chomsky; Lasnik, 2021[1993]). Sendo assim, a observação das razões que permitem marcações diferentes nos parâmetros das línguas são fundamentais para se compreender processos de mudança que interagem para formar as gramáticas particulares das diversas línguas.

A TPP é um desdobramento de estudos iniciados por Chomsky na década de 1950, assim como na década de 1990 surge outro desdobramento, o Programa Minimalista. Segundo Chomsky e Lasnik (2021[1993]), os parâmetros são os responsáveis por determinar os limites de variação entre as gramáticas particulares. Nesse sentido, cada gramática representa um conjunto particular de valores parametrizados, que segundo a teoria gerativa, emergem no processo de aquisição da linguagem do falante a partir da interação entre os princípios da GU e os dados linguísticos a que a criança está exposta.

O Parâmetro do Sujeito Nulo (PSN), também denominado de Parâmetro *pro-drop* (do inglês *pronoun drop*, “queda do pronome”) (Chomsky, 1981, 1986; Perlmutter, 1971; Rizzi, 1982, dentre outros), é o parâmetro mais discutido no escopo da TPP, tendo como seu marco inicial o início dos anos 80. Assim, há a proposição do chamado Princípio da Projeção Estendida (EPP), princípio esse que afirma que todas as línguas possuem sujeito, no entanto, nem todas as línguas o realizam foneticamente. Desse modo, o sujeito nulo é uma interpretação de sujeitos referenciais em sentenças finitas e declarativas nas quais um pronome, um sintagma nominal (SN) ou expressão referencial não estariam realizados foneticamente, com a posição de sujeito permanecendo vazia sem expressão fonológica, conforme exemplo retirado dos dados analisados nesta pesquisa:

Inf.: Minha mãe trabalhava só na roça, Ø plantava muita roça (Informante AR02).

Segundo este parâmetro, uma categoria vazia é permitida na posição (NP, S) nas línguas de tipo *pro-drop* e bloqueada nas de tipo não *pro-drop*. A possibilidade de realização fonética do sujeito pronominal constitui o parâmetro que marca as línguas positivamente ou as marca negativamente, língua não *pro-drop*. As línguas que marcam positivamente o parâmetro de sujeito nulo são chamadas de *pro-drop*. Quando uma língua licencia o sujeito nulo, como é o caso do espanhol e italiano, geralmente apresenta um paradigma rico de concordância verbal de número e pessoa. A partir dos estudos de Huang (1984), a relação direta existente entre riqueza flexional e licenciamento do sujeito nulo deixou de ser determinante, pois o chinês apresenta um sistema sem flexão verbal, no entanto, caracteriza-se por ser uma língua de sujeito nulo. Após este estudo, surgiu uma nova abordagem sobre o parâmetro, a proposta de Jaeggli e Safir (1987), segundo a qual a condição que licencia o sujeito nulo não seria exatamente a riqueza do elemento *AGR*, mas a uniformidade morfológica dos paradigmas verbais de uma língua.

Segundo Chomsky (1981), existe um conjunto de propriedades relacionadas que compõem o Parâmetro *pro-drop*, são eles:

- i) omissão do sujeito;
- ii) inversão livre na ordem do sujeito e verbo;
- iii) extração do sujeito à distância;
- iv) resumptivo nulo do sujeito em orações encaixadas;
- v) ausência do efeito *that-t*.

As análises comparativas entre línguas como o italiano permitiram a constatação de que nem todas as línguas *pro-drop* apresentarão este conjunto de propriedades associadas. Enquanto línguas como o italiano, o espanhol e o PE licenciam o sujeito nulo, outras como o francês e o inglês não o permitiriam. Com o intuito de interpretar o licenciamento do sujeito nulo a partir de uma propriedade morfológica motivada por fatores ligados a essas línguas, alguns pesquisadores dedicaram-se a analisar as particularidades que permitem a omissão do sujeito pronominal, como a riqueza/pobreza flexional das línguas e o modo de orientação (ao discurso ou à sentença).

Para Taraldsen (1978), os tipos de flexão verbal organizariam as línguas do mundo, diferenciando aquelas que permitiriam o apagamento do sujeito daquelas que não permitiriam tal comportamento. Desse modo, Chomsky (1981) afirma que o estabelecimento do parâmetro

pro-drop é o primeiro passo no sentido de buscar explicar diferenças entre línguas no tocante à possibilidade de apresentarem ou não um sujeito nulo. O ponto fundamental seria, portanto, o elemento de concordância – *AGR* – capaz não só de licenciar como também permitir a recuperação do sujeito em línguas com um sistema flexional rico.

A mudança paramétrica, por essas razões, não pode ser concebida de forma instantânea, mas é “caracterizada pela diminuição de frequência de uso de determinada estrutura que passa a ter estatuto gramatical diferente (reanálise diacrônica) à medida que vai sendo substituída por outra” (Duarte, 1995, p. 31). Assim, os dados linguísticos a que a criança tem acesso durante a aquisição não são mais “robustos ou salientes” (Lightfoot, 1991), com o tempo, a tendência é o desaparecimento dessas estruturas, a limitação desses dados a registros estilísticos ou marcadores de construção cristalizadas (Rizzi, 1988). Tem-se neste ponto uma mudança paramétrica, ou seja, a refixação de um novo parâmetro. Duarte (1995, 2018) constatou que o PB passou por uma mudança paramétrica com a perda do princípio “Evite Pronome”, assim, estruturas com sujeito nulo se tornaram cada vez menos frequentes, caracterizada por uma língua de sujeito nulo parcial. À medida que a criança aciona os parâmetros com base em uma evidência positiva, espera-se que cada geração faça reanálises diacrônicas até a refixação do parâmetro.

Silva (2023, p. 39) apresenta a mudança de parâmetro a partir do processo de TLI entre afrodescendentes, conforme o quadro abaixo 2:

Quadro 2 - Ilustração da mudança do PSN durante a aquisição linguística na situação afro-brasileira

	<i>Input</i>	<i>Output</i>
Grupo 1 (colonos portugueses)	+	+
Grupo 2 (escravos africanos)	+	+/-
Grupo 3 (afrodescendente-1)	+/-	+/-
Grupo 4 (afrodescendentes2-n)	+/-	+/- Tendência ao +

Fonte: Silva (2023, p. 39)

O quadro apresentado por Silva (2023) ilustra, o processo de mudança do PSN no contexto da formação do PAF, considerando diferentes gerações envolvidas na situação de contato linguístico. No grupo 1, composto pelos colonos portugueses, observa-se a manutenção

do valor positivo do PSN (+), característica associada ao PE da época, cuja gramática permitia a omissão do sujeito pronominal em virtude de uma morfologia flexional verbal relativamente rica. Já no grupo 2, formado por africanos escravizados que adquiriram o português como segunda língua, verifica-se uma situação de instabilidade no parâmetro (+/-), uma vez que o processo de aprendizagem em condições adversas e de contato intenso entre línguas levou a uma reestruturação parcial da gramática.

Nos grupos subsequentes, constituídos por afrodescendentes nascidos no Brasil, o *input* linguístico passa a refletir essa gramática já parcialmente reestruturada, caracterizada por um sistema variável em relação à realização do sujeito pronominal. Assim, tanto o *input* quanto o *output* passam a apresentar valores mistos (+/-), indicando um estágio de variação e reorganização gramatical que se transmite intergeracionalmente. Esse processo evidencia uma tendência gradual de mudança no comportamento do PSN, resultado das condições históricas de contato linguístico que marcaram a formação do PB, especialmente em contextos afrodescendentes.

Nesse sentido, a representação proposta por Silva (2023) dialoga com as hipóteses da TLI, segundo as quais o contato entre falantes de diferentes línguas, aliado à aquisição imperfeita da língua dominante, provocou simplificações morfossintáticas e reorganizações estruturais na gramática emergente. Tal perspectiva contribui para compreender como a variação no uso do sujeito pronominal no PB, particularmente em variedades populares e afro-brasileiras, pode estar relacionada a processos históricos de contato linguístico e transmissão intergeracional de gramáticas parcialmente reestruturadas.

Assim, de acordo com o quadro 2, em virtude de uma TLI e por influência de sua própria língua, os escravizados adultos, no processo de aquisição do português, produziram dados distintos do PE. Ribeiro (1998, p. 102) destaca que os dados do PE que serviram de *input* para a aquisição da língua portuguesa pelos brasileiros foram produzidos por diferentes gramáticas, a depender da época. A autora destaca duas gramáticas portuguesas que ofereceram dados para a aquisição no Brasil: a dos séculos XVI-XVIII e a dos séculos XIX-XX. Silva (2023, p. 39-40) destaca que apesar de receberem uma marcação positiva para o PSN (*input*) provavelmente a alteraram posteriormente (*output*). Almeida e Carneiro (2009, p. 73) consideram que os africanos e seus descendentes podem ter tido acesso à gramática do sujeito nulo, tanto através da gramática do PB quanto através da gramática do PE (Almeida; Carneiro, 2009, p. 73).

Holmberg e Roberts (2009), ao analisar um grupo distinto de línguas indo-europeias, ampliaram a noção de sistemas mistos de Chomsky e apresentaram algumas línguas que

admitem sujeitos nulos (isto é, alguma propriedade associada ao parâmetro *pro-drop*), classificando-as em quatro grupos:

1. Línguas de sujeito nulo canônicas: línguas que licenciam o sujeito referencial não expresso/nulo, tais como o italiano, o português europeu, o grego, o turco e o basco. Nessas línguas, a pessoa do discurso pode ser retomada através da flexão verbal.
2. Línguas de sujeito nulo parcial: nessas línguas, os sujeitos sem referência definida são, em geral, nulos, havendo restrições na aplicação do parâmetro *pro-drop* em sujeitos referenciais. Exemplos: o finlandês, o hebraico e o português brasileiro.
3. Línguas de sujeito nulo radicais: línguas que licenciam a não expressão tanto o sujeito quanto do objeto, sem que haja flexão verbal indicativa da pessoa do discurso. Como exemplo, tem-se algumas línguas asiáticas, tais como o chinês, o japonês, o coreano, entre outras, cujo licenciamento se dá através do discurso.
4. Línguas de sujeito nulo expletivo: línguas em que um sujeito referencial não pode ser nulo, mas um sujeito expletivo pode. Exemplos: o holandês e o alemão.

Assim, enquanto línguas como o italiano licenciam a não expressão do sujeito através da concordância verbal, línguas como o chinês e o japonês o fazem através dos níveis de referencialidade desses sujeitos (Kato; Duarte, 2014), o que nos leva a concluir que são muitas as perspectivas de abordagem do licenciamento da não expressão do sujeito, sobretudo nas línguas românicas.

3.3 O Parâmetro do sujeito nulo em variedades do português

O PE possuía maior diversidade de desinências verbais associadas para cada pessoa do discurso; ele se caracteriza, assim, por ser uma língua de concordância verbal forte. Diante disso, pode-se afirmar que o PE é uma língua de sujeito nulo consistente, sendo o mesmo recuperável por meio da desinência verbal (Raposo, 1992, p. 478). Assim, a categoria vazia na posição de sujeito no PE é constante, uma vez que o sujeito é recuperado pela morfologia verbal. No entanto, o comportamento dessa variedade não é algo uniforme se comparado as línguas de sujeito nulo radical, uma vez que a interpretação não pode ser garantida apenas pela flexão, mas por um antecedente expresso no contexto sintático, discursivo ou pragmático.

Um estudo que revelou o comportamento do sujeito nulo no PE foi realizado por Duarte (1995) a partir de amostras de fala do português coloquial europeu. Duarte analisou 30 entrevistas gravadas em 1987 com informantes de Portugal. Os informantes da amostra foram divididos em duas faixas etárias (22-33 anos e 46-60 anos) e em três níveis de escolaridade. Duarte (1995) constata que o sujeito nulo é predominante em todas as pessoas do discurso no PE, conforme o quadro 3:

Quadro 3 - Frequência de sujeitos nulos no PE

	Em oração coordenada	Sem oração coordenada
Pessoa	Sujeitos nulos (%)	Sujeitos nulos (%)
1 ^a	60	53
2 ^a	73	72
3 ^a	73	64

Fonte: Duarte (1995, p. 8).

As menores ocorrências de sujeitos nulos são para a primeira pessoa, que apresenta desinência exclusiva tanto para o singular quanto para o plural. Esses dados dão evidências de que na variedade de PE analisada, a riqueza da flexão verbal pode não ser tão relevante para o preenchimento ou não preenchimento da posição do sujeito. Outra conclusão do estudo se refere à frequência em orações coordenadas e não-coordenadas. A diferença na frequência entre os tipos de oração é muito pequena. Nota-se, nesse sentido, que na segunda pessoa, o percentual de sujeitos nulos nas orações não-coordenadas praticamente não é alterado. Duarte (1995) não encontrou, igualmente, diferença significativa quanto ao uso do sujeito nulo e as variáveis sociais, faixa etária e níveis de escolaridade; isso demonstra o estágio avançado de apagamento do sujeito realizado no PE. Outra constatação importante diz respeito ao traço semântico do sujeito, o traço [-animado] favorece o sujeito não realizado, correspondendo a 93% de nulos com o traço [-animado], enquanto 69% dos nulos apresentam o traço [+animado].

Há, todavia, um fato considerado por Duarte (1995) como destoante na regularidade que opera o sujeito nulo no PE: em suas amostras, o sujeito pleno supera o nulo na primeira e terceira pessoas em contextos de relativas, interrogativas indiretas e interrogativas *qu*-. Além desse fato, nos dados foram encontrados apenas 7% de sujeitos nulos associados ao traço [-animado], um

percentual muito reduzido quando se compara com o italiano e o espanhol, cuja recusa ao pronome pessoal é absoluta quando associado a este traço.

A discussão inicial sobre o PSN no PE, no entanto, se diferencia em diversos aspectos do caso brasileiro. As primeiras discussões sobre a marcação paramétrica no PB não encontraram uma definição precisa, pois, essa variedade não se encaixou como uma língua de sujeito nulo e nem como uma língua de sujeito preenchido, sendo a sua interpretação distinta do PE. Jaegli e Safir (1989) propõem que o sujeito nulo é licenciado por uma morfologia uniforme, no entanto, o PB se apresenta como um contra exemplo para esta proposta, pois sujeitos nulos são licenciados mesmo não havendo uniformidade morfológica. Já a proposta de Roberts (1993) para o PSN é que ele envolve dois tipos de riqueza para *AGR*, uma formal e outra funcional. Nessa perspectiva, uma morfologia é formalmente rica se apresenta um morfema para cada pessoa, e será funcionalmente rica se permitir apenas um sincretismo e uma forma zero.

Novamente o PB não se encaixa nesta proposta, pois não apresenta um paradigma nem formalmente nem funcionalmente rico, uma vez que apresenta formas derivadas e não derivadas e possui mais de uma forma zero e mais de um sincretismo, como é o caso da segunda e terceira pessoa. Silva (2004, p. 584) mostrou a inadequação da hipótese da Uniformidade Morfológica proposta por Jaeggli e Safir (1989), tendo por base a análise de Pratas (2004), segundo a qual o licenciamento de expletivos nulos não está associado ao fato de a morfologia ser rica. Ao comparar estruturas frásicas do PB com o PE, o inglês e o francês, Silva (2004, p. 584), apresentou contra-evidências à binaridade do PSN (Rizzi, 1988, p. 1997), pois, segundo ela, o PB caracteriza-se como uma língua de sujeito nulo parcial, tendo como base a proposta da hierarquia paramétrica no âmbito do programa minimalista.

Almeida (2005, p. 13) argumenta que essas especificidades tornam o PB “uma língua singularmente rica para a discussão do parâmetro do sujeito nulo exatamente porque parece não se encaixar nessas formas de licenciamento apresentadas”. A autora acrescenta também a existência de duas grandes discussões referentes aos estudos que tratam do sujeito nulo no português brasileiro. Uma primeira vertente questiona se ainda há sujeito nulo na gramática “ativa” dessa língua ou se são meros “resíduos” de gramáticas anteriores (Tarallo, 1993; Duarte, 1993, 1995). Já a outra é defendida por aqueles que acreditam que esta língua ainda é *pro-drop* (Negrão, 1999; Negrão e Muller, 1996; Negrão e Viotti, 2000), apresentando, segundo esses estudiosos, formas específicas que estariam levando o PB a ser uma língua de sujeito nulo

diferente de outras línguas românicas (Galves, 1984, 1990; Kato, 1999, 2000; Figueiredo Silva, 1996; Modesto, 2000, entre outros).

Duarte (1993) evidencia a relação entre a expressão do pronome e a redução no paradigma flexional do PB a partir da gramaticalização de *você* (que passa a substituir *tu* a partir dos anos 1935) e a entrada da forma *a gente* (competindo com o pronome *nós*, mas facilmente vencido na segunda metade do século XX), conforme quadro 4:

Quadro 4- Evolução nos paradigmas flexionais do PB

	PRONOMES NOMINATIVOS	PARADIGMA 1 SÉCULO XIX	PARADIGMA 2 SÉCULO XX/1	PARADIGMA 3 SÉCULO XX/2
1 os	Eu	Canto	canto	canto
1 PP	nós <i>a gente</i>	Cantamos	cantamos cantaØ	cantamos cantaØ
2 os	tu <i>você</i>	Cantas -	cantas cantaØ	canta(s) cantaØ
2 PP	vós <i>vocês</i>	Cantais Cantam	- cantam	- canta(m)
3 os	ele, ela	cantaØ	cantaØ	cantaØ
3 PP	eles, elas	Cantam	cantam	canta(m)

Fonte: Duarte (2018, p. 85)

O sistema pronominal mudou, segundo a autora, de um sistema com seis formas distintas para um paradigma que apresenta quatro ou cinco formas, devido a inserção de formas nominais que se gramaticalizaram como pronomes – *a gente* e *você*. Ambas as formas se combinam com o verbo na terceira pessoa do singular, conforme o paradigma 2. Nas palavras de Duarte (2018):

Como se vê, o português do Brasil evoluiu de um sistema com seis formas distintas, mais dois sincretismos – representados pela segunda pessoa indireta, que utiliza as formas verbais de terceira pessoa (Paradigma 1) – para um paradigma que apresenta quatro formas, graças à perda da segunda pessoa direta (Paradigma 2). Este paradigma, restrito hoje à língua escrita e à fala de uma geração situada numa faixa etária mais alta, coexiste com um terceiro, em que se veem apenas três formas, em consequência da perda do pronome de primeira pessoa do plural *nós*, substituído na fala dos jovens e, cada vez mais, popular entre os falantes de faixas etárias mais altas, pela expressão *a gente*, que se combina com formas verbais de terceira pessoa do singular (paradigma 3) (Duarte, 2018, p.110).

O sujeito nulo, assim, prevaleceu, no Paradigma 1, com as formas *tu* e *vós*, referindo-se a uma segunda pessoa direta, com uma morfologia própria e distintiva e também às formas de

tratamento que se referem a uma segunda pessoa indireta, que tomam emprestadas as formas verbais da terceira pessoa. Duarte (2018) constatou que o índice de sujeitos nulos foi de 69% em 1918 e de 25% em 1937, coincidindo com a perda da segunda pessoa direta. Segundo os dados de Duarte, a curva descendente teria se iniciado em 1918 e, só na segunda metade do século XX, o uso do sujeito pleno de primeira pessoa superou o do sujeito nulo.

A tese de Duarte (1995) constitui um trabalho muito relevante sobre o PSN no âmbito do PB, principalmente pela profundidade das análises realizadas que serviram de referência para estudos posteriores. Em sua tese, a autora conclui que apesar de uma redução para a primeira pessoa na segunda metade do século XX, o sujeito nulo continua sendo a opção preferida, criando uma assimetria: de um lado, os sujeitos de primeira e segunda pessoas representados com mais frequência pelo pronome lexical e, do outro, o sujeito de terceira pessoa, que usufrui da opção permitida às línguas *pro-drop*. Assim, de acordo com Cyrino, Duarte e Kato (2000), ao analisarem, por um lado, a redução do uso de sujeitos nulos e, por outro, o crescimento do objeto nulo no PB, as autoras observam que a referencialidade exerce papel central na escolha entre formas pronominais. Para as autoras, em línguas que admitem tanto variantes nulas quanto não nulas, um fator decisivo para a seleção de uma ou outra forma é o estatuto referencial do elemento a que se faz referência no discurso, conforme tabela 5.

Tabela 5 - Hierarquia Referencial

não-argumento	proposição	3p. +/-humano +/-específico	2p. 1p. +humano
[-ref] < ----- > [+ref.]			

(Cyrino, Duarte, Kato, 2000, p. 59)

Com base nessa hierarquia, os argumentos que apresentam os traços [+nominal] e [+humano] ocupam o nível mais elevado da escala de referencialidade, enquanto os não-argumentos situam-se na posição inferior. No que se refere aos pronomes pessoais, o falante (eu) e o interlocutor (você), por representarem participantes humanos do ato de fala e corresponderem à primeira e à segunda pessoas, encontram-se no topo dessa hierarquia. Já a terceira pessoa ocupa uma posição mais baixa, em razão da interação entre os traços [+/-humano] e [+/- específico].

Nessa hierarquia, o comportamento do sujeito pode ser compreendido a partir do grau de referencialidade associado a esses diferentes elementos. Em posição ainda mais baixa encontra-se o sujeito que retoma uma proposição, frequentemente denominado sujeito neutro. Por fim, no extremo inferior da hierarquia estão os sujeitos não referenciais, que não apresentam referente discursivo identificável.

A possibilidade de sujeito nulo na terceira pessoa numa língua *pro-drop* está limitada a certas restrições. A autora destaca ainda a atenção dispensada por Calabrese (1986), referente à complementaridade entre o uso do pronome pleno e do pronome nulo da terceira pessoa em italiano:

Quando o referente é esperado, usa-se o pronome nulo; caso contrário, o pronome pleno. E o que faz um referente ser fortemente esperado é o fato de ser ele o sujeito de uma predicação, termo o qual Calabrese prefere referir-se como TEMA, usando uma terminologia tradicional. Este é o referente disponível para um pronome nulo numa matriz ou numa subordinada a ele ligada. Só uma barreira entre o pronome o TEMA rompe a necessária adjacência sintática, forçando o uso do pronome pleno. Embora sem o caráter obrigatório do italiano, isto é o que ocorre nos trechos analisados (Duarte, 1996, p. 116 -117).

A elevação do nível de realização do sujeito pronominal, segundo a pesquisadora, estaria relacionada com a substituição do pronome pessoal *tu* pelo pronome de tratamento *você* e com a concorrência da expressão *a gente* com o pronome *nós*; processos que ganham corpo nas primeiras décadas do século XX e que levam a uma expressiva redução na flexão verbal, pois, tanto com *você*, quanto com *a gente*, o verbo se mantém na forma da terceira pessoa do singular, ou seja, sem um morfema específico de pessoa e número. Para Duarte (1996), o português do Brasil apresentava antes de 1937, características de língua *pro-drop*, mas a redução no quadro de desinências verbais alterou essa condição. Segundo ela, o que ocorreu no português do Brasil, aconteceu também em relação ao francês medieval. O português do Brasil estaria passando por uma fase de transição de língua *pro-drop* para língua não *pro-drop*, através da perda da riqueza funcional do seu paradigma verbal. Fatos como este fazem parte de um conjunto de mudanças sintáticas que nos permitem perceber o desenvolvimento de uma gramática brasileira (Tarallo, 1993).

Ribeiro (1998), ao refletir sobre a formação do PB, numa perspectiva diacrônica a partir das características sintáticas do português que aqui chegou, afirma que o português para cá trazido são dois, pelo menos, tendo em vista o contexto de gramáticas em competição na aquisição do português por brasileiros (Ribeiro, 1998, p. 101). Segundo Ribeiro (1998), para

melhor entender as características do PB, é preciso “procurar as raízes da primeira gramática do PE dos séculos XVI e XVII, não desconsiderando, contudo, influências outras – das línguas africanas e indígenas” (Ribeiro, 1998, p. 116).

Tarallo (1993) já afirmava que mudanças sintáticas dramáticas no português brasileiro no final do século XIX estabeleceram uma nova gramática distinta da de Portugal. Muito embora reconhecesse que essas mudanças haviam se processado muito antes do período indicado, o autor afirma que “as circunstâncias sociais podem não ter sido suficientemente satisfatórias para que a pena brasileira começasse a escorrer sua própria tinta” (p. 99). O estudo de Tarallo constatou, nesse sentido, modificações no paradigma verbal do PB que implicaram em um maior preenchimento da posição de sujeito. A partir do quadro abaixo, é possível perceber uma indicação da mudança na marcação do parâmetro.

Quadro 5 – Retenção pronominal

Tempo/função	1725	1775	1825	1880	1981
Sujeito	23.3%	26.6%	16.4%	32.7%	79.4%
Objeto direto	89.2%	96.2%	83.7%	60.2%	18.2%
(Sps)	96.5%	98.9%	91.3%	72.9%	44.8%

Fonte: Tarallo (1993)

Desse modo, o final do século XIX e início do século XX foram determinantes nesse processo, pois os dados apontam um percentual muito maior de pronomes expressos na posição de sujeito a partir da década de 1880. Do mesmo modo que Tarallo (1993) e Duarte (1995), Berlinck (1989) também defende um processo de transição no qual a língua portuguesa no Brasil passava de um sistema *pro-drop* para não *pro-drop*. Segundo ela, o PB estava deixando de ser uma língua que possui sujeito nulo e que o enfraquecimento da concordância e a mudança do sistema pronominal tiveram um papel relevante para este processo.

A perda de uma propriedade com determinado valor paramétrico, quando relacionada a outra, pode implicar novos rumos que a mudança pode tomar. Com relação à propriedade de inversão livre do sujeito, considerada por Chomsky (1981, p. 254) como a propriedade fundamental das línguas *pro-drop*, foi perdida pelo PB, como comprovam as pesquisas de Berlinck (1989, 1995), Duarte (1992), Lopes Rossi (1993), Ribeiro (1995). Assim, a ordem VS ficou restrita a sentenças apresentativas, o que nos fez perder a estabilidade em relação ao parâmetro e permitiu que outras mudanças pudessem surgir.

Segundo Duarte (2019), o PB exibe um comportamento distinto e diverge da constatação de Galves (1987), Holmberg (2010), Holmberg e Sheehan (2010), cujos estudos convergem para o entendimento do PB como uma língua de sujeito nulo parcial. Para a referida autora, o PB caminha num processo de mudança para uma língua de sujeito preenchido, processo esse já concretizado na primeira e segunda pessoa do discurso, mas ainda mostra resistência em alguns contextos de terceira pessoa, pois a terceira pessoa do singular, a qual discutimos em detalhes no próximo capítulo, possui propriedades peculiares, podendo ser recuperada por correferência a um SN antecedente.

Duarte afirma, assim, que o PB está passando por uma mudança de língua de sujeito nulo para uma língua de sujeito pronominal preenchido, e não necessariamente que é, intrinsecamente, um sistema parcial. Os dados evidenciam a mudança concretizada na primeira e segunda pessoa, sendo o contexto de resistência a terceira pessoa, que, mesmo aumentando a frequência de sujeitos realizados de 25% nos anos 1880 para 58% em 1990, resiste ao preenchimento do sujeito diante do traço semântico [- específico] e a contextos em que o referente se encontra distante do sujeito.

Duarte (2020) constata que importantes propriedades das Línguas de Sujeito Nulo Consistentes (LSNC) foram perdidas pelo PB e que a mudança se encontra em fase final em relação ao PB urbano. Algumas evidências apontadas por ela que explicam esse já encaixamento da mudança é o surgimento de estruturas com o sujeito deslocado à esquerda e retomado por um pronome fraco, o chamado redobro do sujeito, estruturas inexistentes em LSNC:

2. a. *Aí, [muitas dessas pessoas]_i elas_i estudavam artes cênicas.*

b. *Eu acho [que essa área de programação visual]_i ela_i é um pouco mais difícil.*

Outro subproduto da mudança apontada por Duarte (2020) são as alterações nas estratégias de indeterminação do sujeito. Enquanto as LSNC usam o clítico “se” para expressar referência arbitrária, o PB perdeu o clítico “se”, e prefere a terceira pessoa do plural (com ou sem pronome expresso) para expressar a referência arbitrária, reservando o clítico para a escrita:

3. a. *a gente sabe das coisas assim... ah Ø_{arb} matou fulano... e quando Ø_{arb} mata, alguém faz alguma coisa.*

Para Duarte (2020), a única estrutura inesperada num sistema que muda em direção ao preenchimento e prefere indeterminar com pronomes nominativos expressos é o sujeito nulo com o verbo na terceira pessoa do singular:

4. a. $\emptyset_{\text{gen}}$ Tem de ajudar o ser humano a se desenvolver... como ser humano.
- b. se $\emptyset_{\text{gen}}$ quer comprar alguma coisa $\emptyset_{\text{gen}}$ vem pra Copacabana.

Esta possibilidade de um nulo genérico, além dos sujeitos nulos referenciais em encaixadas com *c*-comando, levam Holmberg (2010) e Holmberg e Sheehan (2010) a incluir o PB entre as línguas de sujeito nulo parcial. Entretanto, Duarte (2020) afirma que essas estratégias não estão aumentando entre as faixas mais jovens, o que não sugere mudança em curso, além da constatação de que os nulos referenciais caminham para a extinção e os nulos arbitrários e genéricos são pouco frequentes e podem perfeitamente ser substituídos por formas mais prototípicas, como:

5. a. Na televisão (eles) fala muito isso.
- b. Você tem de ajudar o ser humano a se desenvolver... como ser humano.

Além dessas evidências de mudança, Duarte destaca as sentenças impessoais e a reestruturação dos verbos inacusativos. Para ela a possibilidade de ocorrências como:

6. a. $\emptyset_{\text{expl}}$ Acabou a bateria do meu celular.
- b. A bateria do meu celular acabou.
- c. [Meu celular]_i acabou a bateria [t_i].

Evidenciam que o não desenvolvimento de um expletivo lexical, como o francês medieval, permitiu um alçamento do genitivo (5c) e não de todo o DP (5b), enrijecendo assim a ordem SV em sentenças apresentativas. Assim, “parece que a solução encontrada pelo sistema em mudança foi mover elementos referenciais para Spec, TP. Isso sugere que não apenas a redução do paradigma flexional, mas também a proeminência de tópico devem ser levadas em conta no processo de mudança por que passa o PB” (Duarte, 2020, p. 96).

Em suma, é consenso entre os estudiosos, com exceção de Nicolau (1995), que houve uma mudança paramétrica no PB, mas há divergências no sentido desta mudança. Duarte (2020, p. 95) alerta para o cuidado na inclusão do PB entre as línguas de sujeito nulo parcial, pois esse

rótulo “parcial” pode conter sistemas muito diferentes, exigindo uma análise rigorosa dos contextos que licenciam sujeitos nulos (Biberauer, 2010). Os estudos de Duarte possuem, nesse quadro, grande relevância, pois fornecem bases empíricas para a análise do fenômeno na variedade culta do PB. Diante do processo de polarização que marca a formação do português no Brasil, convém compreendermos como se manifesta este parâmetro no português popular, especificamente na variedade afro-brasileira.

Para esse comparativo, tomaremos como base o estudo de Lucchesi (2004), com dados de fala coletados em três comunidades rurais afrodescendentes e isoladas no interior da Bahia, em Cinzento, Helvécia e Barra/Bananal e o estudo de Almeida (2005) realizado em três comunidades rurais de municípios baianos: Bananal/ Barra dos Negros e Mato Grosso, ambas em Rio de Contas, e Matinha, em Feira de Santana. A análise do licenciamento do Parâmetro *pro-drop* nesta variedade do PB buscou verificar se comunidades geograficamente isoladas e afetadas pelo contato entre línguas africanas e o português exibiriam especificidades quanto a realização do sujeito pronominal.

O estudo de Duarte (1993) forneceu bases importantes para analisar o licenciamento do sujeito nulo na variedade culta do português do Brasil, a partir de dados de fala em peças teatrais do século XIX. No entanto, Lucchesi (2009) pondera que, com relação aos mecanismos de concordância verbal e nominal, a norma popular apresentará uma tendência diferenciada de mudança ou exibirá um estágio mais avançado de perda da propriedade de licenciamento do sujeito referencial nulo, aproximando-se mais das línguas crioulas de base lexical portuguesa da África. Essas questões são imprescindíveis nesta pesquisa, que desenvolve uma análise variacionista da realização do sujeito pronominal de referência definida na fala de uma comunidade rural afro-brasileira.

Com respeito ao PSN em línguas como o PB, Lucchesi (2009) afirma que nessas línguas, “os processos de alteração nos valores dos parâmetros sintáticos constituem a parte mais nebulosa da questão de como as situações de contato afetam a estrutura gramatical da língua” (Lucchesi, 2009, p. 167). Os resultados obtidos, no entanto, foram diferentes do que era esperado, de um total de 4.599 ocorrências, 1.255 dados eram de uso do sujeito nulo, o que corresponde a 27% de sujeitos nulos. Este percentual foi muito próximo do obtido por Duarte (1995), cuja variedade analisada foi o português “culto”. Segundo o autor,

Por outro lado, pode-se pensar, com base na visão das duas grandes vertentes evolutivas do PB proposta por Lucchesi (1994, 1998, 2001, 2002), que houve uma espécie de convergência na evolução gramatical dessas duas variedades linguísticas. Enquanto, na norma culta, a diminuição no nível de sujeito

pronominal nulo derivou de uma reorganização estrutural em função das substituições dos pronomes tu e nós por você e a gente, respectivamente; na norma vernácula, essa diminuição estaria ligada diretamente à queda de morfemas flexionais de pessoa e número do verbo, que não seria o resultado de uma reorganização da pauta dos pronomes pessoais sujeito, mas um reflexo direto do processo de transmissão linguística irregular que teria atuado mais intensamente na formação das variedades populares do PB (Lucchesi, 2009, p. 176-177).

Segundo ele, a razão para o progressivo aumento da realização de pronomes ocupando a posição de sujeito no PAF foi o enfraquecimento da concordância verbal, enquanto no PB “culto”, o sincretismo causado pela entrada dos pronomes *você* e *a gente* foram os agentes que desencadearam a mudança na marcação do parâmetro. Essas variedades convergiram por razões diferentes no sentido de não permitir livremente a ocorrência de sujeitos nulos como na língua-alvo. Com relação à pessoa do sujeito, os dados de Lucchesi revelaram que a taxa mais alta de sujeitos nulos é com ocorrências da terceira pessoa do singular, correspondendo a 34% de sujeitos nulos. Esse resultado, novamente, assemelha-se ao encontrado por Duarte no PB culto. Resta-nos, portanto, compreender o comportamento desse fenômeno na comunidade quilombola baiana Baixa da Linha.

4 PRESSUPOSTOS TEÓRICOS

[...] pode-se conceber o quadro de profunda variação na concordância nominal e verbal que se observa hoje nas variedades populares do português do Brasil como uma decorrência do processo de transmissão linguística irregular que marca a gênese dessas variedades linguísticas. (Lucchesi, 2009, p. 124).

Este capítulo é dedicado a apresentar as bases teóricas que sustentam este estudo. Inicialmente, na subseção 2.1, apresentaremos alguns conceitos basilares da Sociolinguística Variacionista, discutindo aspectos desde a sua gênese teórica aos principais conceitos estabelecidos. Nos itens 2.1.1 e 2.1.2, desenvolvemos o conceito de variável linguística e debatemos os cinco problemas da mudança linguística, respectivamente. Na subseção 2.2, discutimos sobre o contato linguístico e sobre como este reflete, sobretudo, nas variedades populares do português do Brasil, dando ênfase ao processo de transmissão linguística irregular, no item 2.2.1, e ao processo de histórico de formação do português brasileiro, marcado pela polarização sociolinguística, o qual é apresentado no item 2.2.2. No item 2.3 realizamos algumas considerações sobre o contexto sócio-histórico da Bahia, a partir da sua ocupação demográfica e outros fatores históricos. Na última subseção, 2.4, apresentamos algumas evidências sobre o contingente de africanos escravizados trazidos para o Recôncavo, assim como uma breve caracterização da comunidade analisada neste estudo.

4.1 A Sociolinguística Variacionista

Segundo Meillet (1906), as tentativas de se explicar a mudança linguística no século XIX não progrediram devido ao foco das explicações sobre os fenômenos das línguas operarem no sentido da proposição de leis que consideravam apenas as possibilidades de mudança e não as necessidades de desenvolvimento das línguas. Para o autor, a forma mais poderosa de se explicar a mudança linguística viria de uma análise das condições sociais. Dessa forma, Meillet assume posição contrária as ideias vigentes no século XIX, a exemplo do que propunha Kurylowicz (1960), que era totalmente contrário à consideração de fatores externos à estrutura linguística, assumindo uma visão radical sobre os aspectos sociais na análise linguística: “É

preciso explicar os fatos linguísticos com outros fatos linguísticos, não com fatos heterogêneos. [...] A explicação por meio de fatos sociais é um descarrilamento metodológico” (Kurylowicz [1959] *apud* Pulgram (1961, p. 324).

Para Weinreich, no entanto, que retomou, junto a Labov e Herzog, diversos estudos da tradição histórico-comparativa, “a forma mais poderosa de explicação da mudança linguística viria de uma análise das condições sociais” (Weinreich; Labov; Herzog, 2006 [1968], p. 114). Embora Meillet (1906), um dos primeiros a apontar essa necessidade, procurasse se aventurar na análise de fatores sociais, seu domínio limitou-se à mudança lexical. Outros estudos históricos de domínio da linguagem como os de Wyld (1936), Joos (1952), Kökeritz (1953) e Fónagy (1956) percorreram profundamente o contexto social da mudança histórica, no entanto, essas contribuições permaneceram periféricas no desenvolvimento de uma teoria linguística organizada e cumulativa (Weinreich; Labov; Herzog, 2006 [1968], p. 115).

Meillet (1906, p. 17), que já reivindicava em favor do papel de fatores sociais na mudança linguística, afirmou que a língua

[...] é uma instituição com autonomia própria; [...] mas como a língua é também uma instituição social, e o único elemento variável ao qual se pode apelar a fim de explicar a mudança linguística é a mudança social, da qual as variações linguísticas são somente as consequências [...] (Meillet, 1906, p. 17).

Assim, para os autores, as mudanças sociais são as responsáveis por promover as mudanças linguísticas. Nesta perspectiva, o caminho traçado por Weinreich, Labov e Herzog (WLH) no desenvolvimento de sua teoria assume que a compreensão da língua não está no indivíduo, mas nas comunidades de fala. Labov (1972), nesse sentido, definiu-a como a participação do indivíduo em um conjunto de normas compartilhadas; que podem ser observadas em tipos evidentes de comportamento avaliativo e pela uniformidade de padrões abstratos de variação (Labov, 1972 [2008]). Segundo Labov, 1972 [2008], a comunidade de fala é o conjunto de indivíduos que interagem verbalmente e compartilham um conjunto de normas relativo aos usos. Ele propôs, assim, a primeira definição de comunidade de fala, enfatizando a produção linguística, percepção social e avaliação. Portanto, Labov rompe com o pressuposto que considera a língua como propriedade do indivíduo – homogeneidade – discordando de grandes correntes linguísticas como o gerativismo.

Em suma, o modelo inovador de WLH reside na concepção de língua como um sistema heterogêneo ordenado, o qual é condição *sine qua non* para o estudo da mudança linguística. O

termo “ordenada” permite “atribuir à variação um caráter sistemático e controlado que até então lhe fora negado” (Weinreich, Labov e Herzog, 2006 [1968], p. 133). O trabalho do linguista consiste, para os autores, em não apenas reconhecer a variação, mas explicá-la, identificar os fatores que a controlam, inseri-la dentro de um modelo de linguagem, compreender os padrões que governam esta sistematicidade presente em todos os níveis linguísticos em todas as línguas naturais.

Para Labov (1972 [2008], p. 241), em linhas com seus estudos pioneiros, o principal objetivo da Sociolinguística Variacionista é “o estudo da língua em seu contexto social”, correlacionando, assim, fatores sociais e internos à língua com a variação linguística. A utilização do termo “sociolinguística¹³” por si só é rotulado por Labov, 1972 [2008] como um termo “estranhamente redundante” (Labov, 1972 [2008], p. 215). Segundo ele, a língua é uma forma de comportamento social, sendo usada por seres humanos em um contexto social para comunicar ideias, necessidades e emoções uns aos outros. Acrescenta que “sociologia da linguagem” seria o rótulo mais adequado para se tratar desta área de pesquisa. Bright (1966) argumenta que

De fato, é correto afirmar que os estudos sociolinguísticos, da mesma forma dos abarcados sobre o nome ‘sociologia da linguagem’, tratam da relação entre língua e sociedade. Mas tal afirmação é excessivamente vaga. Se tentarmos ser mais exatos, observaremos que a sociolinguística difere de algumas preocupações anteriores com as relações língua-sociedade, pelo fato de que, seguindo novas perspectivas da própria linguística, considera tanto a língua quanto a sociedade como sendo uma estrutura e não uma coleção de itens. A tarefa da sociolinguística é, portanto, demonstrar covariação sistemática das variações linguística e social, e, talvez, até mesmo demonstrar uma relação causal em uma ou outra direção (Bright, 1966, p. 17).

Assim, para além dos questionamentos de nomenclatura, o principal objetivo da Sociolinguística Variacionista, também conhecida como Sociolinguística Quantitativa (Labov, 1972; Weinreich; Labov; Herzog, 1968), é analisar a correlação entre a variação sistemática das línguas e a estrutura das sociedades. Para tanto, esta corrente procura explicar os mecanismos que integram o surgimento das variantes linguísticas, compreendendo que a heterogeneidade e a variabilidade são inerentes à competência linguística dos falantes. Para Lucchesi (2015), é a

¹³ A primeira proposta de uso do termo sociolinguística para trabalhos que relacionam a estrutura social e as formas linguísticas foi realizada em 1952, por Haver Currie, no trabalho “A projection of socio-linguistics: the relationship of speech to social status”. Esse termo, no entanto, só se popularizou nos anos 1960. [Referência: Haver C. Currie (1952) A projection of socio-linguistics: The relationship of speech to social status, *The Southern Speech Journal*, 18:1, 28-37, DOI: 10.1080/10417945209371247]

heterogeneidade da língua que garante a sua funcionalidade em uma comunidade socialmente estratificada e culturalmente diversa. É a flexibilidade conferida pela variação linguística que permite a uma mesma língua funcionar tanto nas feiras livres quanto nas sessões dos tribunais de justiça. Se fosse um código monolítico e inflexível, como sugerem os puristas, a língua não poderia funcionar em ambientes culturais tão diversos, o que levaria fatalmente à sua divisão e fragmentação (Lucchesi, 2015, p. 17).

Tal como a sociedade abriga a diversidade cultural, não se admite que no aspecto linguístico não ocorra variação, uma vez que a língua garante interação entre os seus usuários e está sujeita às coerções e necessidades dessa interação. Afinal, como apontam Weinreich, Labov e Herzog, a língua está a serviço de uma comunidade complexa e, portanto, o que na verdade seria disfuncional “é a ausência da heterogeneidade estruturada” (Weinreich; Labov; Herzog, 2006 [1968], p. 36). Assim, a língua é concebida pelos sociolinguistas como dotada de “heterogeneidade sistemática”, pois é o fator de identificação entre grupos sociais.

A capacidade dos falantes em “alternar códigos no meio de um enunciado, de uma frase ou de uma oração, não uma vez mas várias vezes” (Weinreich; Labov; Herzog, 2006 [1968], p. 105) demonstra a complexidade da estrutura do sistema linguístico. Um dos principais axiomas da Sociolinguística Variacionista, como desenvolvido anteriormente, é a heterogeneidade ordenada. É este aspecto desta disciplina científica que demonstra que a língua varia, e que esta variação é controlada por fatores diversos, tanto estruturais quanto sociais. A sociolinguística concebe as línguas pela sua heterogeneidade, pois comporta, nas palavras de Pagotto (2006), “no seu interior formas em variação e, é o fato de comportar variação que faz com que a língua seja capaz de expressar, no próprio interior do sistema em si, a estrutura social, valores sociais que a ele são externos” (Pagotto, 2006, p. 54). A pesquisa sociolinguística deve partir sempre, portanto, da interrelação entre uma dimensão linguística e outra extralinguística, uma vez que as alternâncias de usos linguísticos são motivadas ou por fatores estruturais ou por fatores sociais. Segundo Tagliamonte (2006),

a Sociolinguística Variacionista é mais apropriadamente descrita como o ramo da Linguística que estuda as características principais da linguagem em equilíbrio entre si - estrutura linguística e estrutura social; significado gramatical e significado social - aquelas propriedades da linguagem que requerem referência a fatores externos (sociais) e internos (sistêmicos) em sua explicação (tradução nossa) (Tagliamonte, 2006, p. 5).

Portanto, para a sociolinguística, o componente social não pode ser desprezado ao se analisar o funcionamento do sistema linguístico. Torna-se fundamental o olhar para a comunidade e não apenas ao falante de modo individual. Os estudos de Labov em Martha's Vineyard¹⁴ e na comunidade de fala de Nova York¹⁵, nesse sentido, foram fundamentais para postular alguns princípios sociolinguísticos acerca das relações de variação estilística, estratificação social e avaliação subjetiva. Meillet (1921, p. 16) observou que todas as leis da história linguística até então descobertas eram meras possibilidades, pois faltava apresentar as condições que permitiam ou provocavam as possibilidades reconhecidas. Identificar os fatores empíricos que condicionam a mudança continuava a ser o problema. Labov se concentrou, em tais estudos, nos mecanismos da mudança linguística. Sua linha de investigação é baseada na doutrina uniformitarista, pois, “os mesmos mecanismos que operaram para produzir as mudanças em larga escala do passado podem ser observadas em ação nas mudanças que presentemente ocorrem à nossa volta” (Labov, 1994 [2008], p. 192).

4.1.1 As variáveis linguísticas dentro do sistema

O estudo da variação linguística começa com a introdução do conceito adicional de variável linguística. Labov desenvolveu este conceito fundamental para analisar as formas inovadoras tão comuns a todas as línguas. Para ele, os fenômenos linguísticos são guiados pela possibilidade de sofrer variação e mudança – *regra variável*, ou pela impossibilidade da variação, pois tal uso poderia implicar na plena comunicação entre falantes - chamada *regra categórica*, pois nem todos os fenômenos linguísticos são passíveis de variação.

Ao comparar o crioulo jamaicano com o inglês padrão, Labov argumenta que “os falantes podem alternar códigos no meio de um enunciado, de uma frase ou de uma oração” (Weinreich; Labov; Herzog, 2006 [1968], p. 105). O autor desenvolve esse novo conceito

¹⁴ A dissertação intitulada *The Social History of a Sound Change on the Island of Martha's Vineryard* (trad. A história social de uma mudança sonora na ilha de Martha's Vineyard), apresentada em 1963, é um marco inicial para o desenvolvimento da Sociolinguística Variacionista.

¹⁵ Para a construção da sua tese, Labov partiu para a análise de dados, os quais coletou em 1962 em três lojas de departamento na cidade de Nova York (Labov, 1972 [2008], p. 64). A pesquisa, conhecida como *The Social Stratification of English in New York City* (trad. A estratificação social do inglês em Nova Iorque), teve como objetivo testar a estratificação social dos vendedores. Labov buscou verificar se a variável linguística representada pela consoante [r] em posição pós-vocálica, quando pronunciada pelos falantes, poderia ser "um diferenciador social em todos os níveis da fala" da cidade, uma vez que, ao não se pronunciar o /-r/, revelava-se uma marca do falar nova-iorquino; logo, pronunciando-o, isso poderia representar uma marca de diferenciação social.

dentro do modelo de heterogeneidade ordenada para dar conta dessa variação íntima. Esse conceito consiste em “um elemento variável dentro do sistema controlado por uma única regra” (Weinreich; Labov; Herzog, 2006 [1968], p. 105). A variável linguística, portanto, difere dos elementos de variação livre na distribuição de suas variantes, já que podem ser identificados fatores condicionantes que, mesmo não sendo estritamente linguísticos, são de interesse linguístico. Este conceito é uma tentativa de definir qual é esse interesse. Assim, uma variável linguística precisa ser definida sob condições precisas para integrar a estrutura linguística, do contrário, haverá apenas a apresentação de regras imprecisas quanto a sua aplicação. Para Labov (2008):

A definição de uma variável linguística é o primeiro e também o último passo na análise da variação. Começa com o simples ato de perceber uma variação – que existem duas maneiras alternativas de dizer a mesma coisa. Uma primeira tentativa de definir envelope de variação busca o maior ambiente em que essa variação ocorre, a fim de aplicar o princípio da responsabilização: que os relatórios das ocorrências de uma variante devem ser acompanhados por relatórios de todas as não ocorrências. A definição de uma variável linguística requer então o estabelecimento de um conjunto fechado ao qual se aplicam os axiomas da teoria das probabilidades. Definições mais precisas da variável são alcançadas localizando e deixando de lado as neutralizações: ambientes em que não é possível distinguir todas as variantes da variável (tradução nossa) (Labov, 2008, p. 1).

O ponto de vista da Sociolinguística Variacionista é que formas linguísticas diferentes, porém equivalentes, concorrem entre si, não de modo acidental, pois há motivações internas e externas à estrutura linguística que permitem a variação. Convém destacar que “nem toda variabilidade e heterogeneidade na estrutura linguística implica mudança; mas toda mudança implica variabilidade e heterogeneidade” (Weinreich, Labov e Herzog, 2006 [1968], p. 125).

A noção de variável linguística foi fundamentada a partir da variação constatada no nível fonológico nas pesquisas desenvolvidas por Labov em Martha’s Vineyard e em Nova York. Críticas ao conceito criado por Labov restringiram a possibilidade de aplicação desse conceito para outros níveis de análise linguística. Estudos como o de Lavandera (1978) questionaram esse modelo pela dificuldade de se tratar a equivalência semântica devido à dificuldade de delimitar quais as variantes relacionadas a determinadas variáveis sintáticas e morfossintáticas. O que se põe em questão, como possível resposta, é que as variáveis linguísticas não podem ser preenchidas por formas de mesmo significado, mas por estruturas que apresentam o mesmo valor de verdade ou até a mesma função quando em contextos equivalentes (Freitag, 2009).

Os falantes atribuem, e isso é o mais importante, diferentes valorações sociais às variantes que compõem uma determinada variável. Quanto à avaliação social, Labov (2008) propõe uma divisão das variáveis em três categorias: os *estereótipos*, que são traços linguísticos socialmente marcados de forma consciente pelos falantes, são formas linguísticas caracterizadas pela extrema estigmatização entre os falantes. Já os *marcadores* são traços linguísticos sociais e estilísticos que permitem um julgamento consciente ou inconsciente do ouvinte sobre o falante. Como exemplo de marcador, sobretudo em falantes cultos, é a variação no uso do clítico acusativo em que há variação na fala mais coloquial (eu encontrei ele) e na fala mais monitorada (eu vou encontrá-lo / eu encontrei-o). Os marcadores são variantes valorizados, o que ocorre nesse caso é a variação estilística, a depender do grau de monitoramento do falante. Os *indicadores*, por sua vez, são os traços socialmente estratificados que não estão no nível de consciência dos falantes, pois não envolve variação estilística. Na base social a concordância pode ser um marcador ou indicador, este é o caso da variável analisada neste estudo, pois o uso variável do sujeito nulo e do sujeito preenchido não é perceptível para os falantes, ou seja, não há um processo de monitoramento, logo não ocorre variação estilística. O modo como uma variável linguística é avaliada socialmente por um indivíduo ou grupo é fundamental para que se possa prever a forma como uma mudança linguística poderá vir a acontecer em uma dada comunidade. É possível, assim, falar em mudanças “de baixo para cima”, em que variantes inicialmente estigmatizadas passam, por serem características de comunidades mais amplas, a dominar o sistema, ou mudanças “de cima para baixo”, em que formas ditas de prestígio passam a ser copiadas pelos grupos mais amplos de falantes, geralmente de classes baixas e mais numerosas.

O conceito de variável linguística é, portanto, fundamental para compreendermos um dos fenômenos variáveis no PB, a realização do sujeito pronominal, cujo comportamento em uma comunidade remanescente de quilombo no interior da Bahia, se constitui em objeto de investigação nesta pesquisa.

4.1.2 As cinco questões da mudança linguística

Como destacado na seção anterior, nem todos os fenômenos linguísticos variáveis conduzem à mudança. Esta é vista como “um tipo de variação linguística com propriedades sociais particulares” (Chambers, 1995, p. 349). É justamente devido a esta estreita relação entre variação e mudança que permite ao pesquisador captar o processo contínuo e gradativo da

mudança, já que a sociolinguística variacionista entende que o processo de mudança não é um processo catastrófico (i. é imediato). A Sociolinguística Variacionista usa, portanto, de ferramentas estatísticas para avaliar se, em um determinado fenômeno, há ou não movimento de mudança a favor de uma ou outra variante por meio do tempo aparente. Quando não há, tem-se uma *variação estável*, que pode ser observada, “quando jovens e velhos apresentam o mesmo comportamento linguístico, e esse comportamento se contrasta com o exibido pela população de meia idade, principalmente pela população que estiver no mercado de trabalho” (Coelho et al, 2015, p. 88), uma vez que este grupo costuma usar uma linguagem mais monitorada e mais próxima das variedades cultas. Quando uma determinada variante se sobrepõe às outras, tem-se uma *mudança em curso*. Esse processo implica competição entre duas ou mais formas variantes, momento no qual a forma inovadora é menos recorrente, e, à medida em que os contextos de uso vão se expandindo, a frequência dessa forma aumenta até que ela ultrapasse a forma antiga (Coelho et al, 2015, p. 73-74). É possível citar como exemplo, segundo os estudos de Duarte (1995, 2018, 2020), a realização do sujeito pronominal, uma mudança em curso que caminha a passos largos rumo ao preenchimento da posição de sujeito.

Weinreich, Labov e Herzog, 2006 [1968] desenvolveram, assim, uma teoria de mudança capaz de explicar como a estrutura linguística de uma determinada comunidade de fala se modifica no curso do tempo, no entanto, tanto a língua quanto a comunidade permanecem ordenadas. Encontrar explicações consistentes para a mudança é a tarefa do pesquisador. Para este trabalho desafiador, a sociolinguística fornece ferramentas para guiar uma pesquisa desta natureza. Essas ferramentas se relacionam a um conjunto de problemas empíricos para os quais o pesquisador precisa fornecer respostas em uma pesquisa sociolinguística. Discutiremos brevemente as cinco questões norteadoras.

1. O problema da restrição (ing. *constraints problem*)

Este problema busca identificar o conjunto das mudanças linguísticas possíveis e as condições para que essas mudanças ocorram numa estrutura linguística. A restrição envolve, pois, a busca por princípios universais que possam prever as direções da mudança. Segundo Coelho et al (2015) “esses princípios não são absolutos, aplicados categoricamente, mas apontam para regularidades e tendências gerais” (Coelho et al, 2015, p. 77).

Inicialmente, são os fenômenos linguísticos particulares que devem ser observados como condições motivadoras de variação e mudança numa dada língua. Entretanto, os

fenômenos linguísticos não podem ser estudados isoladamente, mas correlacionados a outros fenômenos tanto intra quanto interlinguísticos. A este ponto, Lucchesi (2012, p. 795) questiona uma restrição estritamente “universal”, pois esta consiste na busca por uma faculdade da linguagem isolada, que não está encaixada na matriz mais ampla da estrutura linguística e social. Assim, o autor destaca que o “problema das restrições” proposto por Weinreich, Labov e Herzog (1968) estava equivocado e que o problema das restrições deveria ser fundido com o problema do encaixamento.

2. O problema da transição (ing. *transition problem*)

Este problema procura compreender como as formas em variação/mudança se propagam, passando de um determinado estágio para outro a partir da expansão dos contextos linguísticos de uso das formas, pela sua transmissão geracional, pela difusão ao longo do tempo e entre grupos sociais. Assim, a transição verifica qual o percurso da nova forma a partir da identificação dos estágios intermediários da transição de uma forma A para B.

3. O problema do encaixamento (ing. *embedding problem*)

O problema do encaixamento diz respeito ao modo como um fenômeno em mudança se insere no sistema sem que este precise mudar por completo e sem que este se desarticule. Em outras palavras, esse problema deriva do fato de que o sistema continua estruturado enquanto muda. O problema do encaixamento visa, portanto, avaliar a correlação entre a estrutura linguística e social e os usos de determinada variável linguística sem que, com isso, fatores como a comunicabilidade e a eficiência do sistema sejam prejudicados. O encaixamento é toda correlação de fatores e contextos que irão impulsionar ou retrainir a mudança.

4. O problema da avaliação (ing. *evaluation problem*)

A questão da avaliação diz respeito à atitude subjetiva e consciente dos falantes em relação às formas linguísticas em variação ou mudança. Portanto, reações negativas podem retardar ou até impedir a mudança linguística, à medida que os falantes se identificam com tais usos e, rejeitando-os, passam a evitar formas inovadoras. Através da avaliação da consciência

social sobre o processo de variação e mudança, é possível fazer uma previsão do futuro de determinada forma linguística numa comunidade de fala.

Apesar da variação ser o foco da Sociolinguística, o "contato entre as línguas, questões relativas ao surgimento e extinção linguística, multilinguismo, [...] e mudança constituem temas de investigação na área" da Sociolinguística (Mollica, 2013, p. 10), ou seja, ela preocupa-se com assuntos que cruzam língua e sociedade. Portanto, para além da análise da variação, convém investigar as implicações do contato entre línguas que muito pode contribuir para elucidar processos de variação e mudança linguística, em especial por serem esses processos alguns dos gatilhos para a construção de estereótipos. Essa questão será focalizada na próxima seção.

5. O problema da implementação (ing. *actuation problem*)

A mudança linguística pode abarcar motivações ou restrições na sociedade e na estrutura da língua. A implementação, por fim, envolve a investigação dos fatores que podem contribuir para a efetivação de novas formas linguísticas no sistema, além de explicar os contextos linguísticos ou os lugares em que ela ocorre. Em resumo, a questão da implementação visa entender como a estrutura linguística de uma comunidade se transforma no curso do tempo. Sendo assim, explicações mais consistentes sobre a implementação irão requerer um estudo em tempo real, ou seja, quando a mudança estiver em processo e for finalizada.

Como exemplo em relação à implementação na estrutura linguística, o fenômeno variável da realização do sujeito pronominal, que é objeto de investigação nesta pesquisa, tem como característica que a implementação da variante inovadora, o sujeito expresso, depende de fatores linguísticos como o traço [+/- humano] e a acessibilidade do referente. Considera-se, portanto, que o ponto de partida para a implementação do preenchimento do sujeito no português brasileiro sejam os contextos de traço [+ humano] do referente e de menor acessibilidade ao referente. Já com relação à implementação na estrutura social é preciso levar em consideração fatores como identidade, configuração na hierarquia social, percentual demográfico, pois numa comunidade de fala, grupos de prestígio podem acelerar uma mudança em relação aos grupos que se integram à comunidade.

4.2 O CONTATO LINGUÍSTICO

Segundo Lucchesi (2015), grande parte da população que passou a utilizar o português no Brasil colonial não teve acesso pleno ao modelo da língua europeia, aprendendo-a em condições irregulares de transmissão. O ponto crucial sobre a formação da realidade sociolinguística brasileira reside na importância ou não do contato entre a língua do colonizador com as diversas línguas faladas por indígenas e africanos. Diante das diferentes posições assumidas, Araújo (2010) sintetizou três posicionamentos presentes até o início dos estudos da Sociolinguística no Brasil: 1) a supervalorização do contato entre línguas, assumida por Raimundo (1933) e Mendonça (1933); 2) a negação de uma possível influência do contato entre línguas na constituição do PB, conforme Silva Neto (1963) e Chaves de Melo (1972) e; 3) a disseminação de que a presença maciça de africanos ao longo da sócio-história do PB apenas acelerou tendências de mudança já prefiguradas no sistema, segundo Câmara Jr. (1972).

Como não se sabe com precisão quando os primeiros africanos escravizados chegaram ao Brasil, Prado Jr. (1974) faz referência à sua vinda em 1532 desde a primeira exposição de povoadores. A partir de 1549, o tráfico negreiro para o Brasil é autorizado, aumentando a importação de escravizados africanos, principalmente em Pernambuco e na Bahia, onde, “já no final do século XVI, os africanos ocupavam majoritariamente a base da sociedade colonial brasileira; situação que iria se acentuar no século XVII” (Mendonça, 1933, p.32-33). Nesse contexto, Lucchesi (2004, p. 26) afirma que “pode-se pensar que as reflexões sobre as alterações estruturais produzidas pelo contato entre línguas não devem se restringir às situações inequívocas de pidginização e criouliização”, mas a outras alterações observadas na variação na concordância nominal e verbal, aumento do uso de pronomes sujeitos explícitos e reorganização de estruturas sintáticas nas variedades da língua.

O fenômeno da criouliização está diretamente relacionado à expansão do colonialismo europeu entre os séculos XVI e XIX pelos quatro continentes. Segundo Lucchesi (2019), para que a criouliização aconteça, é preciso que haja uma ruptura radical na transmissão da língua de um povo para outro no sentido do dominador para o dominado dentro da estrutura social colonial, de modo que, para o autor, “essa ruptura decorre do fato de a grande massa de escravos ter um acesso muito restrito à língua do grupo dominante que eles são obrigados a adquirir e a usar” (Lucchesi, 2019, p. 3). Assim, a aquisição da língua do grupo dominante se dá a partir de um vocabulário restrito que permite ao grupo dominado construir frases desprovidas de elementos gramaticais. Com o tempo, os povos dominados reestruturam esse código de

comunicação emergencial a partir de sua língua nativa (língua de substrato). A crioulistica denomina este processo de pidginização. No entanto, esta língua pidgin formada pode se tornar nativa para uma comunidade, o que dá origem, portanto, a uma língua crioula.

Embora haja dificuldade na definição de língua crioula entre alguns autores (Egraff, 2001; Mcwhorter, 2001; Chaudenson, 2003; Ansaldo; Matthews, 2007; Bakker; Davalmarkussen; Parkvall; Plag, 2011), Lucchesi (2019, p. 244) afirma que “os crioulos são línguas qualitativamente distintas de suas línguas lexificadoras (LL), e não uma mera variedade da LL, como defendido por Chaudenson (2003), DeGraff (2003) e Mufwene (2001), entre outros crioulistas”. Para ele, essas línguas representam “o poder adaptativo e criativo da faculdade da linguagem humana” (Lucchesi, 2019, p. 250), muito embora os falantes dessas línguas sejam vistos como inferiores devido à sua incapacidade de falar “corretamente” a língua europeia, a qual goza de elevado prestígio.

Um dos estudos sobre essa questão no tocante do PB, já guiado pela sociolinguística, foi o de Guy (1981). Nessa tese, o autor postulou que o português popular do Brasil (PPB) seria derivado de um crioulo português, iniciado entre os séculos XVI e XVII. Essa afirmação, no entanto, não tem sido mais aceita, em virtude da ausência de registros históricos irrefutáveis (Naro; Scherre, 2007) e pela ausência de uma variedade linguística de natureza crioula atual ou que historicamente apontem para a indubitável existência de marcas crioulizantes¹⁶ no PPB (Parkvall; Lópes, 2003). Dentre os linguistas que se posicionam a favor da teoria do contato, há aqueles que defendem que houve a formação de crioulos típicos no Brasil. Neste trabalho, no entanto, partimos do ponto de vista defendido por Lucchesi (1999, 2001, 2003, 2009), de que é possível ter ocorrido a formação de alguns crioulos, porém foram situações isoladas e que não justificariam a atual configuração do PPB. Sobre o contexto da formação do PB, Lucchesi (2009) argumenta que:

O contato dos colonizadores portugueses com milhões de aloglotas, falantes de mais de mil línguas indígenas autóctones e de cerca de duzentas línguas que vieram na boca de cerca de quatro milhões de africanos trazidos para o país como escravos, é, sem sombra de dúvida, o principal parâmetro histórico

¹⁶ A comunidade afro-brasileira de Helvécia, situada no Extremo Sul do Estado da Bahia, é a única comunidade para a qual se têm evidências de que pode ter passado por um processo pretérito de crioulização (FERREIRA, 1984). Porém, segundo Zimmermann (1999), é um caso isolado e com algumas particularidades, como o fato de ter sido oriunda de grandes propriedades de colonos suíços e alemães, que se instalaram na região, no século XIX. Alguns estudos linguísticos sobre Helvécia podem ser encontrados em Baxter (1992, 1996, 1998, 1999), Holm (1992), Megenney (1993), Baxter e Lucchesi (1993 e 1997), Baxter, Lucchesi e Guimarães (1997) e Zimmermann (1999).

para a contextualização das mudanças linguísticas que afetaram o português brasileiro (Lucchesi, 2009, p. 41).

Nessa perspectiva, o contato linguístico constitui um fator fundamental para compreender a formação histórica do PB. A convivência entre colonizadores portugueses, populações indígenas e milhões de africanos escravizados criou um ambiente sociolinguístico caracterizado por intenso multilinguismo, no qual o português foi adquirido como segunda língua. Esse contexto favoreceu processos de simplificação e reorganização estrutural, contribuindo para mudanças linguísticas que diferenciam o PB de outras variedades da língua (Lucchesi, 2009). Petter (2009) destaca que a formação de diferentes variedades do português fora da Europa deve ser compreendida no contexto do contato linguístico estabelecido entre o português e diversas línguas africanas durante o período colonial. Esse contato contribuiu para a emergência de um *continuum* afro-atlântico do português, no qual se inserem variedades africanas, crioulas e afro-brasileiras da língua.

Assim, os negros trazidos da África e os indígenas são considerados elementos fundamentais para a compreensão da realidade social e linguística brasileira. Demograficamente, africanos, descendentes e índios aculturados eram maioria, enquanto portugueses e seus descendentes integravam menos de um terço da população (Lobo, 1996). Assim, os africanos tiveram papel central na formação do chamado “português geral brasileiro”, antecessor do português popular. Segundo Cunha (2003), a formação de variedades populares rurais do PB deve ser compreendida à luz de processos históricos de contato linguístico, nos quais a transmissão do português ocorreu de forma irregular, contribuindo para a simplificação de determinados aspectos da morfologia e da sintaxe nessas variedades

Lucchesi e Baxter desenvolveram, como solução possível para o PPB, o conceito de TLI pois, segundo eles, o PPB seria o resultado de um processo de TLI mais leve (Lucchesi, 2008; 2012; 2015, p. 95-112; Lucchesi; Baxter, 2009), distinto de um processo de criouliização em que a intensidade da TLI é muito mais radical e profunda.

4.2.1 A Transmissão Linguística Irregular (TLI)

As situações de aquisição de linguagem nem sempre são prototípicas, uma vez que os indivíduos necessitam da exposição de dados linguísticos de falantes adultos para configurar a Língua I (Chomsky; Lasnick, 2021[1993]).

O modelo do bioprograma, proposto por Bickerton (1981), parte da ideia de que os seres humanos possuem um mecanismo inato para aquisição da linguagem. Em contextos de contato linguístico intenso, como plantações coloniais, onde falantes de diferentes línguas precisavam se comunicar, surgiam *pidgins*, que possuíam estrutura gramatical reduzida. Quando crianças passavam a adquirir esse *pidgin* como língua materna, elas ativariam esse “bioprograma”, reorganizando o sistema linguístico e produzindo uma língua crioula com estrutura gramatical mais completa. Desse modo, a transmissão linguística é considerada irregular, pois a língua não é transmitida de forma plena de uma geração para outra. Em vez disso, podem ocorrer redução da morfologia flexional; simplificação de paradigmas gramaticais; regularização de padrões sintáticos e criação de novos mecanismos gramaticais para compensar lacunas estruturais. Em crioulos de base portuguesa, como o cabo-verdiano, observa-se a ausência de flexão verbal de pessoa, conforme os exemplos de Holm (1988) e Veiga (1995) abaixo:

N **ta papia** (eu falo)

Bo **ta papia** (tu falas)

El **ta papia** (ele fala)

Nota-se em crioulos de base portuguesa, como o cabo-verdiano, a ausência de flexão verbal de pessoa, sendo o tempo ou aspecto frequentemente marcado por partículas, como em *N ta papia*, *Bo ta papia*, *El ta papia* (Holm, 1988; Veiga, 1995). Devido à forte redução da morfologia flexional verbal, é difícil identificar o sujeito, o que torna necessário a realização do pronome explícito, o qual desempenha um papel central na estrutura da frase.

No caso do PB, o conceito de transmissão linguística irregular (Baxter, Lucchesi, 1997; Lucchesi, 1999 e 2003; e Lucchesi; Baxter, 2009) tem se constituído como um modelo mais amplo de análise da aquisição para além das situações típicas de contato intenso entre línguas, principalmente a aquisição no bilinguismo. Segundo Lucchesi e Souza (2018):

A concepção da transmissão linguística irregular como um conceito gradual visa a desenvolver um modelo mais amplo de análise das mudanças linguísticas induzidas pelas situações de contato maciço entre línguas que não se restrinja às situações típicas de criouliização (Baxter; Lucchesi, 1997; Lucchesi, 2002, 2008; Lucchesi; Baxter, 2009). A ideia básica é que uma situação de contato entre línguas massivo pode conduzir à formação de uma língua crioula, que tem uma gramática qualitativamente distinta da língua alvo (Rouge, 2008), mas também pode resultar na formação de variedades históricas da língua dominante com algumas características estruturais das línguas crioulas, porém em um nível menos intenso e mais superficial,

caracterizando uma transmissão linguística irregular de tipo leve (Lucchesi; Souza, 2018, p. 467).

Desse modo, ao passo que uma “TLI radical gera uma nova língua, qualitativamente distinta da LL, uma TLI mais leve gera apenas uma variedade da língua dominante” (Lucchesi, 2019, p. 246). O conceito de TLI refere-se, portanto, a um quadro teórico amplo e gradual destinado a explicar diferentes configurações de contato entre línguas. Tal noção abrange desde situações mais intensas, típicas de processos de criouliização, nas quais emergem variedades linguísticas estruturalmente diferenciadas da língua dominante no contexto de contato, até contextos menos radicais. Nestes últimos, como no caso das variedades populares do PB, o contato entre línguas desencadeou mudanças estruturais na gramática da língua, sem, contudo, produzir uma ruptura suficientemente profunda que resulte na constituição de uma gramática qualitativamente distinta daquela do português.

A questão que se coloca, como consequência, é a necessidade de reestruturação gramatical provocada pelas situações de contato maciço entre línguas, perdidas numa fase inicial do processo de aquisição precária da língua alvo por falantes adultos.

Estudos sobre o contato entre o português e línguas africanas mostram que diferentes variedades do português formadas na África e no Brasil apresentam características estruturais convergentes, permitindo falar em um *continuum* afro-brasileiro do português (Petter, 2009). Esse cenário histórico favoreceu processos de TLI, que contribuíram para a reestruturação de determinados aspectos da gramática dessas variedades.

Lucchesi (2009, p. 167) apresenta três situações de recomposição da estrutural gramatical que atuam sobre a variedade linguística que se forma: (i) na redução da morfologia flexional; (ii) na alteração da marcação de parâmetros sintáticos em função de parâmetros não marcados; e (iii) na gramaticalização de itens lexicais para substituir os elementos gramaticais perdidos no estágio inicial do contato.

O panorama sociolinguístico do Brasil revela que quanto maior o nível do contato na formação de uma variedade linguística, menor a frequência de uso da morfologia flexional quanto às regras de concordância. Assim, o português que se configura no continente americano possui especificidades fruto da intensidade do contato sobretudo entre africanos e portugueses. Como bem disse A. Houaiss (1985, p. 91), o “português brasileiro nasce com diversidade”, pois a “língua geral brasileira” a partir do século XVIII dava continuidade a duas possibilidades: “um português africanizado ou um português europeizado” (Mattos e Silva, p. 21), resultando em perfis diferenciados ao se considerar o conjunto brasileiro. Como ocorreram “diversas

situações sociolinguísticas e variados graus de contacto”, para Castro (1980, p. 13), a busca por evidências históricas que se proponham a reconstruir sobre a face negra do português brasileiro constitui um trabalho que pode levantar hipóteses sobre o passado.

4.2.2 A polarização sociolinguística do Brasil

O Brasil desenvolveu duas grandes vertentes linguísticas: variedades mais próximas do PE, associadas às elites letradas e variedades populares, formadas em contextos de contato linguístico (Lucchesi, 2015). Desse modo, a realidade social e linguística do Brasil é bipolarizada, de um lado, os falantes da(s) norma(s) culta(s) e de outro os falantes da(s) norma(s) popular(es). Tal contexto, observado desde o período colonial brasileiro, proporcionou a formação de uma variedade de língua modificada devido ao contato com outras línguas. Essa variedade de língua se mantém e pode ser observada facilmente entre falantes das camadas populares, tanto urbanas quanto rurais, como é o caso da CRQ aqui estudada. Convém destacar aqui que o falante da norma popular brasileira é aquele com pouca ou nenhuma instrução, analfabetos ou semianalfabetos, os quais correspondem a mais de 60% do total da população rural (Lucchesi, 2015, p. 158). Esta pesquisa busca descrever a realidade linguística desta parcela da população que se mantém na extremidade oposta à realidade da elite culta do Brasil.

O que define a realidade social da língua no Brasil é o que Lucchesi (2015) chama de polarização sociolinguística. O Brasil, entre 1532 e meados de 1695 se caracterizava por um *multilinguismo generalizado* (Mattos e Silva, 2004). O português, língua minoritária da elite colonial, era apenas uma das línguas faladas no território, pois convivia com centenas de línguas indígenas e africanas faladas por um grande contingente populacional. Enquanto em São Paulo a maioria dos filhos de mulheres indígenas com colonizadores portugueses (chamados de mamelucos) falavam a língua geral paulista, a qual era uma versão alterada do tupi; no atual estado do Maranhão e do Pará, a língua geral era outra, de base tupinambá, que predominou na antiga província do Grão-Pará¹⁷. Na Bahia e em Pernambuco, impulsionado pela crescente importação de escravizados africanos para o trabalho nos engenhos de cana-de-açúcar, houve um cenário para o desenvolvimento de uma língua banto, com base nas línguas bantu africanas⁸.

¹⁷ Ainda hoje esta língua é falada no Brasil, na região da Amazônia, cujo nome é *nheengatu*, que traduzido é “língua boa” (RODRIGUES, 1993, 2006). ⁸ Principalmente o quimbundo.

Assim, Lucchesi (2015, p. 86) argumenta que é provável que variedades pidginizadas e crioulizadas de português podem ter convivido com uma língua africana franca, baseada no quimbundo (Silva Neto, 1963 [1951], p. 85).

Apesar deste cenário de multilinguismo, o avanço da língua portuguesa entre os séculos XVII e início do século XIX foi desencadeado pela chegada de colonos portugueses ao Brasil em busca do enriquecimento que as minas de ouro despertaram e pelo grande contingente de africanos sequestrados para atender ao trabalho da mineração, que tiveram o português imposto por diversos decretos coloniais. A estimativa é de que ao menos 300 mil colonos portugueses tenham vindo ao Brasil, enquanto a média anual de importação de escravizados passou de 7 mil indivíduos, nas últimas décadas do século XVII, para 15 mil ao longo do século XVIII (IBGE, 2000, p. 223). Diante desse contexto demográfico, no Sudeste, a língua portuguesa avança sobre a língua geral paulista e posteriormente também sobre a província do Grão Pará, com o decreto em 1758 do governo português. Lucchesi (2006) denomina esse período de avanço do português como a *primeira grande onda de lusitanização* do território brasileiro. Este período é marcado pela convivência do português com as línguas africanas e variedades pidginizadas e crioulizadas da língua portuguesa, sobretudo nos quilombos (Lucchesi, 2015).

A *segunda onda de lusitanização* (Lucchesi, 2006) teve início com a vinda da Corte portuguesa para o Brasil em 1808, e o consequente processo de urbanização, no Rio de Janeiro decorrente da instalação de membros da elite portuguesa. O fim do tráfico negreiro em 1850 foi um grande marco para a eliminação “da grande fonte do multilinguismo no Brasil, já que a população indígena continuava a ser dizimada e se refugiava nos confins do território brasileiro” (Lucchesi, 2015, p. 87). Um fator que contribuiu para a aspiração de uma cultura europeia no Brasil e, assim, para a maior imposição da língua, por meio de um projeto de “branqueamento” do país (Lucchesi, 2009), foi a imigração europeia e asiática, pois uma vez que o tráfico negreiro fora proibido, tornava-se necessário a importação de mão-de-obra para atuar nas lavouras.

O início do século XX caracterizava-se pelo estado avançado de homogeneização linguística do Brasil. A presença de línguas africanas era residual, com destaque para Salvador, onde o iorubá (língua africana da família *kwa*) ainda subsistia entre os afrodescendentes (Rodrigues, 2004 [1933]). Contudo, Lucchesi (2015, p. 88) destaca que “essa homogeneização linguística não resultou de uma assimilação dos contingentes da base da pirâmide social ao mercado consumidor e ao espaço da cidadania, mas da imposição da língua dominante em um processo de violenta opressão simbólica e cultural”. Assim, Lucchesi (2015) defende uma

mudança no sentido da polarização sociolinguística do Brasil, antes relacionado ao português da elite colonizadora *versus* as línguas indígenas e africanas, para uma situação muito próxima de *diglossia*.

O cenário demográfico do Brasil estava, nesse período, em maior parte ocupado por descendentes de africanos e indígenas, correspondendo a 60% da população, que habitava predominantemente em áreas rurais. Sem acesso à escolarização, mais de 75% dos brasileiros eram analfabetos. Este quadro é resultante da ausência de políticas governamentais, a partir de 1888, que pudessem integrar um contingente populacional pós-escravidão à educação e à geração de renda. Como resultado desse processo de exclusão, um contingente de afrodescendentes foi condenado à pobreza e à marginalidade (Lucchesi, 2015). Lucchesi (2015, p. 88) considera que essa população analfabeta e residente no interior do país “devia falar variedades bastante alteradas do português”. A diferença para o outro extremo, o da elite brasileira, se intensificava no plano ideológico à medida em que estes importavam deliberadamente os modelos linguísticos da antiga metrópole portuguesa para intensificar as diferenças do idioma e separar, também por meio da língua, mestiços e negros, contribuindo para aprofundar a exclusão e o racismo. No entanto, a partir do início do século XX, dois processos atuaram na força dessa polarização, nas palavras de Lucchesi (2015):

A intensidade dessa polarização sociolinguística [...] começou a diminuir de forma mais efetiva, a partir da Revolução de 1930, que pôs fim à República Velha e desencadeou o processo de industrialização e urbanização da sociedade brasileira. Esse vertiginoso processo inverteu, ao longo do século XX, a distribuição da população brasileira no binômio *campo-cidade*, já que, neste limiar do século XXI, mais de 80% da população brasileira vive em centros urbanos [...] (Lucchesi, 2015, p. 88-89).

Desse modo, a polarização sociolinguística começa a diminuir a partir da industrialização e urbanização do país que contribui, mesmo que lentamente, para o acesso à educação das classes econômicas mais baixas. Esses dois processos contribuíram, assim como ocorreu em outras sociedades modernas, para um *nivelamento linguístico*, no qual “a variedade urbana de base letrada se impõe sobre o mosaico de dialetos rurais, [...] sendo a universalização do letramento em seus níveis mais elevados, um poderoso agente desse nivelamento linguístico” (Lucchesi, 2015, p. 89). O modelo de norma urbana culta passa a promover mudanças “de cima para baixo” na norma linguística dos segmentos populares, os quais passam a eliminar as marcas na fala de formas e estruturas resultantes do contato entre línguas.

Contudo, esse processo de nivelamento linguístico não atingiu o nível esperado, devido a fatores econômicos fundados na superexploração do trabalho e na concentração de renda em minorias. Assim, “a polarização sociolinguística do Brasil ainda mantém sua radicalidade, sobretudo no plano ideológico da avaliação social das variantes linguísticas” (Lucchesi, 2015, p. 89). O estigma contra as formas de linguagem popular, que são o reflexo direto das mudanças induzidas pelo contato na formação histórica do PB, revelam o caráter racista do preconceito linguístico, que se constitui como um plano de dominação ideológica arquitetado pela elite brasileira. Esta incursão teórica sobre o contato linguístico e suas implicações na polarização do português brasileiro é fundamental para se compreender as raízes do preconceito linguístico que permanece latente na sociedade brasileira.

4.3 O CONTEXTO SÓCIO-HISTÓRICO DA BAHIA

Reconstruir a história do português brasileiro envolve uma análise abrangente de diferentes fontes linguísticas, históricas e socioculturais. Esse processo é um dos grandes desafios pelo qual se debruçam os linguistas (Castilho, 1984, 2004; Scherre, 2006, 2010; Lucchesi, 1999, 2005, 2017; Martins, 2010, 2015; Mattos e Silva, 2004). A visão que se projeta neste estudo é a de que o contato linguístico, a aquisição da língua portuguesa como segunda língua e o processo irregular de transmissão da língua portuguesa se constituíram como fatos centrais para a atual configuração do PB. Esses fatos, conjugados à história social e demográfica do Brasil, permitiram definir, sociolinguisticamente, um quadro de mudanças que individualizaram o português brasileiro. Segundo Lobo (2001, p. 17), a difusão das mudanças para a quase totalidade dos falantes do português como língua materna, inclusive entre os falantes dos estratos socioeconômicos privilegiados, deu-se em virtude da pressão demográfica exercida pela população proveniente dos estratos socioeconômicos não privilegiados.

4.3.1 A pressão demográfica e a escolarização no Brasil

Os fatores sociais interagem de modo indissociável da dinâmica linguística de uma comunidade, pois, segundo Labov (2008), só há um tipo de linguística, a social. Esses fatores externos à língua interagiram de maneiras complexas para moldar a variedade linguística do português brasileiro. Conforme Mattos e Silva (2004), para que se compreenda e interprete a

heterogeneidade dialetal brasileira – tanto em relação às variedades popular, quanto em relação às variedades cultas - é preciso traçar interrelações com os “fatos das sucessivas distribuições demográfico-linguísticas” e com “a penetração da língua escrita no Brasil” (Houaiss). Referindo-se ao trabalho de A. Mussa ao traçar a história demográfica do Brasil, Mattos e Silva (2004, p. 34), afirma que “a história demográfica fornece pistas interessantes e significativas para a história das línguas”, pois, segundo ela, não se pode compreender a história das línguas sem levar em conta, em igualdade de condições, “os aloglotos¹⁸”, com seus percursos históricos e suas línguas que conviveram e convivem com a língua portuguesa. O quadro abaixo, extraído de Mussa (1991, p. 163), fornece uma tentativa aproximada de reconstrução da sócio-história linguística do Brasil.

Quadro 6: Demografia histórica da sociedade brasileira

	1538-1600	1601-1700	1701-1800	1801-1850	1851-1890
Africanos	20%	30%	20%	12%	2%
negros brasileiros	-	20%	21%	19%	13%
Mulatos	-	10%	19%	34%	42%
brancos brasileiros	-	5%	10%	17%	24%
Europeus	30%	25%	22%	14%	17%
índios integrados	50%	10%	8%	4%	2%

Fonte: Mussa (1991)

A partir da análise do quadro, é possível inferir que durante toda a história brasileira, a maioria foi não branca, o que significa que a língua familiar não era a portuguesa. Essa tabela aponta um perfil demográfico geral da história do Brasil. Esses dados, no entanto, não são suficientes, pois não dão conta, nas palavras de Houaiss (1985, p. 32), “das sucessivas distribuições demográfico-linguísticas” ao longo da história e dos diversificados espaços geográficos brasileiros ao longo dos cinco séculos.

¹⁸ Falantes não nativos do português.

Os dados reunidos por Mattos e Silva a partir de fontes historiográficas primárias e interpretação de historiadores, fornecem informações valiosas sobre a dinâmica linguística do Brasil, a partir da demografia. Esses dados, portanto, representam um passo importante para a compreensão da afirmação de A. Houais (1985, p. 91) de que “o português brasileiro nasce com diversidade”. Isso fica evidente ao observar que, ao final do século XVI, fins do século XVIII e início do século XIX, a taxa de falantes não nativos do português varia entre 63% e 88% da população, enquanto os de língua materna portuguesa variam entre 12% e 37%.

Embora em alguns estudos sobre o português do Brasil prevalecessem o ideal homogeneizador que procurou enobrecer o português brasileiro ancorado à ideia de normatização linguística, não há como negar a difusão e implantação do português no território brasileiro; no entanto, um português distinto do europeu. Os estudos demonstram que, no Brasil, se usavam mais de mil línguas autóctones, de diversos grupos linguísticos. Um desses estudiosos, Aryon Rodrigues (1993, p. 91), propõe que no início do período colonial eram faladas 1.175 línguas indígenas, sendo que 85% delas foram dizimadas, e continuam a desaparecer, assim como seus falantes, correspondendo hoje a cerca de 262.000 falantes (Gomes, 1993, p. 63-64) de cerca de 180 línguas (Rodrigues, 1993, p. 92). Em relação às línguas africanas que aqui chegaram, o grupo banto foi o que demonstrou as taxas mais altas e constantes (35%, 65%, 64%, 50%), seguidos dos benue-kwa não banto (7%, 10%, 20%, 24%), além de outros falantes dos grupos linguísticos do oeste-atlântico, como o mande, kru, gru, dentre outros (Mussa, 1991, p. 144-145).

A essa configuração, incrementa-se a informação sobre a imigração para o Brasil, ocorrida no século XIX, em que “portugueses, sobretudo açorianos, europeus da Europa Central e do Leste europeu, além de asiáticos do oriente médio e do extremo oriente” (Mattos e Silva, 2004, p. 38) chegaram ao Brasil com suas línguas e culturas. Portanto, para superar interpretações simplificadoras como a da “vitória” da língua portuguesa e mesmo teorias conservadoras, é preciso considerar o “tecido multilíngue do Brasil” (Mattos e Silva, 2004, p. 38). Outro fator atuante para a constituição do PB foi o reduzido índice de escolarização que será discutido na subseção seguinte.

4.3.1.2 O processo de escolarização na Bahia

A história do Brasil é marcada pela violência, exploração e desrespeito às diferenças culturais da maior parte da população. Segundo Antunes (2003, p.165), a política educacional envolve os princípios gerais que definem a finalidade da formação escolar, no sentido de se determinar o perfil de pessoa que se espera ter socialmente. Assim, o projeto de colonização português fundava-se no desenvolvimento de um trabalho educativo e missionário que visava catequizar e instruir os índios para que fossem explorados como escravos.

A educação jesuítica, implantada no período colonial, estava integrada à política colonizadora e, por mais de dois séculos, foi o modelo de educação exclusivo no Período Colonial (1500-1822). O processo colonizador dos séculos XVI e XVII “construiu um ‘tupi jesuítico’ - a língua geral da costa, de base tupi, chegou a ser um risco para a hegemonia do português no Brasil” (Mattos e Silva, 2004, p. 14). Além dessa, tanto no litoral quanto nas entradas paulistas, enquanto no Nordeste e na Amazônia se desenvolveram as línguas gerais de base cariri (Houaiss, 1985, p. 49-50) e outra tupinambá¹⁹, respectivamente.

Entretanto, esse modelo de educação, antes centrado na catequização dos indígenas, transformou-se em uma educação de elite, e conseqüentemente, num instrumento de ascensão social. O ensino tornava-se inútil para a grande massa pobre, uma vez que a fonte econômica estava fundada na agricultura e no trabalho escravo, de modo que o ensino não teria utilidade dentro desse sistema de sobrevivência. A Companhia de Jesus, portanto, passa a formar a elite colonial, enquanto as classes mais baixas, incluindo indígenas e africanos escravizados, ficavam à eternamente à espera da instrução. Enquanto isso, o multilinguismo se alastrava.

Com a expulsão dos Jesuítas de todas as colônias portuguesas, em 1759, por ordem do Marquês de Pombal, “a educação brasileira (...) vivenciou uma grande ruptura histórica num processo já implantado e consolidado como modelo educacional” (Bello, 1992). A reforma pombalina dos estudos menores²⁰ objetivou, de acordo com Laert Ramos de Carvalho, “criar a escola útil aos fins do Estado e, nesse sentido, ao invés de preconizarem uma política de difusão

¹⁹ Esta língua é o antepassado do nheengatu, língua brasileira falada no rio Negro.

²⁰ Ensino das primeiras letras. Porém não se limitavam a isso: além do curso elementar, eles mantinham os cursos de Letras e Filosofia, considerados secundários, e o curso de Teologia e Ciências Sagradas, de nível superior, para a formação de sacerdotes. No curso de Letras estudava-se Gramática Latina, Humanidades e Retórica; no curso de Filosofia estudava-se Lógica, Metafísica, Moral, Matemática e Ciências Físicas e Naturais. Os que pretendiam seguir as profissões liberais iam estudar na Europa, na Universidade de Coimbra, em Portugal, a mais famosa no campo das ciências jurídicas e teológicas, e na Universidade de Montpellier, na França, a mais procurada na área de medicina (Bello, 1992, p. 2).

intensa e extensa do trabalho escolar, pretenderam os homens de Pombal organizar a escola que, antes de servir aos interesses da fé, servisse aos imperativos da Coroa” (Piletti, 1992, p. 36).

Outra grande força atuante no cenário linguístico brasileiro, sobretudo no século XIX, foi o processo de normativização linguística, marcado pela relusitanização do Rio de Janeiro, com a transferência da capital do reino em 1808, trazendo ao Brasil alguns milhares de portugueses. A necessidade de escolarizar o povo brasileiro fica evidente na primeira constituição, a de 1824. Os percentuais apresentados por Houaiss (1985, p. 137) demonstram o salto no processo de escolarização, pois aponta que o número de letrados - os que concluem o primeiro grau - no Brasil até o final do século XVIII não ultrapassavam 0,5%, chegando a 30% em 1920.

A sócio-história linguística do português brasileiro está fundamentada na proposição de que a sua gramática é diferente do PE, e que essa definição de gramática se deu na virada do século XIX para o XX, conforme Tarallo (1993, p. 69-106), o que parece ser condizente com os dados históricos para a escolarização massiva e a dizimação das outras expressões linguísticas.

A primeira escola básica fundada na Bahia foi a "Escola Régia de Primeiras Letras da Bahia", estabelecida em 1551 pelo governador-geral do Brasil, Tomé de Sousa. Esta escola foi uma das primeiras instituições educacionais do Brasil colonial e desempenhou um papel fundamental na educação dos filhos dos colonos portugueses e na catequização dos povos indígenas. A Escola Régia de Primeiras Letras da Bahia foi precursora do sistema educacional brasileiro e contribuiu para o desenvolvimento da educação na região e no país.

5 ASPECTOS METODOLÓGICOS

A metodologia da Teoria da Variação constitui uma ferramenta poderosa e segura que pode ser usada para o estudo de qualquer fenômeno variável nos diversos níveis e variações linguísticas. As suas limitações são as do próprio linguista, a quem cabe a responsabilidade de descobrir quais são os fatores relevantes, de levantar e codificar os dados empíricos corretamente, e, de interpretar os resultados numéricos dentro de uma visão teórica da língua (Naro, 1994, p. 24).

A fidedignidade dos resultados em uma pesquisa científica está condicionada à utilização de procedimentos metodológicos precisos e bem definidos. O pesquisador que se debruça sobre dados obtidos por meio de entrevistas sociolinguísticas deve utilizar uma metodologia consistente a fim de fornecer uma observação e interpretação sistemática dos dados, subsidiada pelo pressuposto teórico definido.

Portanto, esta seção visa apresentar o caminho metodológico percorrido, desde a constituição das amostras de fala, com o fim de alcançar os objetivos propostos nesta pesquisa e confirmar ou refutar as hipóteses levantadas. Assim, na subseção 5.1, descrevemos as amostras de fala tal como foram constituídas neste trabalho; logo após, no item 5.2, o envelope da variação; e, por último, no item 5.3, o processamento quantitativo dos dados.

5.1 A CONSTITUIÇÃO DAS AMOSTRAS DE FALA

Antes de iniciar a pesquisa, o projeto precisou seguir os procedimentos estabelecidos pelo Comitê de Ética em Pesquisa (CEP-UEFS), tal rigor constitui etapa primordial e imprescindível para a proteção dos direitos, integridade e dignidade dos participantes da pesquisa²¹. Uma amostra de fala sociolinguística precisa ser estratificada segundo as características sociais dos informantes. Essas informações são organizadas em células sociais de acordo com a metodologia da Sociolinguística de Primeira Onda²² (Labov, 2008[1972]). Um

²¹ Esta pesquisa foi submetida ao Comitê de Ética em Pesquisa (CEP) da Universidade Estadual de Feira de Santana (UEFS) e aprovada sob o parecer nº 7.763.709 e CAAE nº 87057825.5.0000.0053, conforme as normas da Resolução nº 466/2012 do Conselho Nacional de Saúde.

²² Eckert (2005, 2012) propõe uma divisão das práticas dos estudos sociolinguísticos em três “ondas” que, segundo ele não devem ser vistas como estágios que se sucederam ao longo da história sociolinguística, mas correspondem

dos aspectos importantes da constituição das amostras é a representatividade. No entanto, de acordo com Coelho *et al* (2015),

as pesquisas sociolinguísticas têm mostrado que não há necessidade de amostras tão grandes como as usadas em outras pesquisas de natureza social (de intenções de voto, por exemplo) para se analisar fenômenos variáveis, uma vez que o uso linguístico é mais homogêneo do que o comportamento humano acerca de outros fatos (COELHO et al, 2015, p. 114).

Portanto, uma boa amostra não significa ser numericamente saliente, contendo uma quantidade expressiva de informantes, uma vez que os usos linguísticos atuam de modo distinto aos demais fatos sociais, o que fornece uma visão com caráter homogêneo do comportamento linguístico de uma determinada comunidade.

Outro elemento importante da coleta de dados é a confiabilidade. Nesse ponto reside a possibilidade de replicação do estudo, pois, se aplicados os mesmos procedimentos na mesma comunidade, espera-se encontrar resultados semelhantes.

As amostras que constituem o *corpus* dessa pesquisa são compostas de entrevistas sociolinguísticas concedidas por moradores que nasceram na comunidade Baixa da Linha, situada na zona rural da cidade de Cruz das Almas – BA. A análise das amostras de fala possibilita o levantamento de dados estatísticos fundamentais para a comprovação ou não das hipóteses levantadas.

5.1.1 Coleta de dados orais

Os métodos de coleta de dados para constituição de *corpora* oral em pesquisas sociolinguísticas envolvem o contato do pesquisador com indivíduos ou comunidades. O cuidado do linguista neste processo deve focalizar a minimização de danos aos sujeitos

à maneira com que esses modelos lidaram/lidam com a variação linguística ao longo das décadas de estudo. Segundo Coelho et al (2015), os estudos voltados para a primeira onda compreendem a variação visando estabelecer correlações amplas entre categorias sociais e variáveis linguísticas. Já os estudos de segunda onda, utilizam métodos etnográficos para investigar de forma ampla as categorias sociais envolvidas na variação. Por fim, os estudos de terceira onda percebem a variação como um sistema complexo de significados sociais em potencial dentro de uma comunidade, que pode ser manipulado localmente pelos indivíduos na construção de estilos identitários. Embora essas perspectivas sigam caminhos distintos, não são mutuamente excludentes, o que não impede que um mesmo fenômeno seja analisado sob a “ótica da comunidade de fala, que permite um olhar mais amplo e panorâmico sobre o objeto, e, num momento posterior, ser analisado em redes sociais e/ou comunidades de prática, permitindo uma visão mais detalhada, mais "micro" do fenômeno (Coelho et al, 2015, p. 70).

envolvidos na pesquisa. Portanto, buscou-se neste estudo manter preservada a identidade dos participantes.

A não observação das técnicas de coleta parte da noção, conforme apontado por Vaux e Cooper (1999), de que é o dado de fala mais importante do que a qualidade do registro. Para a gravação das entrevistas foi utilizado um gravador LG smartphone.

O *corpus* desta pesquisa é composto por 909 dados extraídos de 12 entrevistas concedidas por falantes de uma variedade afro-brasileira do português, coletadas no ano de 2025, na cidade de Cruz das Almas, Bahia, Brasil, na comunidade quilombola Baixa da Linha. Algumas estratégias de minimizar os ruídos, conforme Oliveira (2014), foram adotadas, como: (I) nas proximidades: que inclui conversas com outras pessoas da família, presença de animais e execução de atividades domésticas; (II) nas redondezas: o tráfego de pessoas e veículos na rua; (III) máquinas: geladeiras, computadores, celulares; (IV) eventos imprevisíveis: manusear de papéis, cadeiras; (V) eventos previsíveis (que podem ser evitados): *feedback* do pesquisador, tamborilar em mesa; (VI) manuseio do equipamento: cabos, microfone, celular. No entanto, há situações em que os ruídos são inevitáveis, e, por se tratar de uma comunidade rural, ruídos inesperados de animais, crianças e transeuntes ocorreram durante as gravações. No entanto, como bem observa Bower (2008), uma gravação feita bem distante do ideal é melhor do que nenhuma gravação. Isto significa que as condições de coleta de dados nunca serão perfeitas, mas com planejamento cuidadoso da recolha de dados, “é possível lançar mão do melhor que se puder” (Oliveira, 2014).

O grupo de fatores condicionantes adotado neste estudo foi desenvolvido no âmbito do projeto Vertentes do Português Popular do Estado da Bahia²³, coordenado e implantado pelo professor Dante Lucchesi no Departamento de Letras Vernáculas da UFBA, em 2002, e que muito tem contribuído para o estudo da realidade sociolinguística da Bahia.

5.1.2 O perfil dos informantes

Os informantes desta pesquisa nasceram na comunidade e/ou cresceram nela. Como buscamos ter acesso a uma versão da língua que tenha sofrido pouca ou nenhuma interferência do processo de normativização imposta pela escolarização formal, buscamos por informantes com pouca ou nenhuma escolaridade. Assim, os informantes foram estratificados de acordo com os critérios comuns adotados nas pesquisas sociolinguísticas: sexo, idade, escolaridade,

²³ <http://www.vertentes.ufba.br/>

estada fora da comunidade. Consideramos três faixas etárias: a faixa I, composta por pessoas de 25 a 35 anos, a faixa II, por pessoas de 45 a 55 anos e a faixa etária 3, formada por informantes a partir de 65 anos. No quadro 7, apresentamos o perfil dos informantes:

Quadro 7 – Perfil dos participantes da pesquisa

CÓDIGO	SEXO	FAIXA	ESCOLARIDADE	OCUPAÇÃO	PERÍODOS FORA DA COMUNIDADE
DN	F	1	semianalfabeta	lavradora	-
LC	F	1	semianalfabeta	lavradora	-
RM	M	1	analfabeto	lavrador	-
RS	F	2	analfabeta	lavradora	-
MP	F	2	analfabeta	lavradora	-
CS	M	2	semianalfabeto	lavrador	-
AR	M	2	analfabeto	lavrador	-
MC	F	3	analfabeta	lavradora	-
VR	F	3	analfabeta	lavradora	-
SS	M	3	analfabeto	lavrador	-

Fonte: elaborado pela autora

Conforme o quadro 7, é possível perceber que todos os informantes exercem atividades agrárias das quais retiram a sua subsistência e a de seus familiares. Em geral, é no interior da comunidade rural que eles passam o maior tempo. Ressalta-se que um dos informantes inicialmente incluídos no *corpus* foi posteriormente excluído da análise, uma vez que não nasceu na comunidade investigada. Considerando que um dos critérios metodológicos desta pesquisa é que todos os participantes sejam naturais da comunidade, optou-se por retirar esse informante do conjunto de dados, a fim de garantir maior homogeneidade sociolinguística da amostra e maior confiabilidade nos resultados obtidos. Além disso, encontramos apenas um informante da faixa etária 1 que fosse analfabeto/semianalfabeto, logo o total de informantes analisados nesta pesquisa foi de 10 (dez).

Os aspectos socioculturais dos informantes são semelhantes, embora em apenas algumas casas os moradores tenham acesso à Internet, mas estão expostos à cultura midiática principalmente por meio do rádio e TV. Algumas casas são de taipa²⁴, mas a maioria das famílias conseguiu substituir a frágil construção por outra de alvenaria.

5.1.3 A natureza das entrevistas

Os contatos com a comunidade se iniciaram em 2023, no entanto a coleta de dados aconteceu em 2025²⁵. As entrevistas tiveram uma média de 40 a 60 minutos, variando a duração conforme cada participante. Todos os informantes foram entrevistados em suas casas, em cômodo de sua escolha, cuidado que teve o intuito de quebrar qualquer tipo de formalidade que representasse um obstáculo à fala espontânea dos informantes. Assim, algumas entrevistas foram gravadas enquanto nos sentávamos em pedras, outras na cozinha.

Os temas tratados na entrevista versaram sobre a vida na roça, as experiências de vida, práticas culturais da comunidade e o trabalho da lavoura. Como o objetivo da entrevista sociolinguística é obter o vernáculo, esses temas possibilitam a espontaneidade da fala, suavizando assim o paradoxo do observador. Segundo Labov (2008, p. 244), “o objetivo da pesquisa linguística na comunidade deve ser descobrir como as pessoas falam quando não estão sendo observadas – no entanto, só podemos obter tais dados por meio da observação sistemática”, motivo que justificou nossa precaução quanto ao conforto subjetivo do entrevistado.

Outra técnica utilizada para reduzir este efeito foi o contato com os participantes antes da coleta de dados. Este contato propiciou maior familiaridade com a comunidade, além de permitir uma entrevista diálogo sobre variados temas que perpassam a história e a cultura da comunidade e região.

As entrevistas gravadas na comunidade Baixa da Linha foram armazenadas em meio digital e posteriormente passaram por um processo de transcrição grafemática, segundo critérios elaborados pelo Projeto Vertentes²⁶, cuja pontuação utilizada indica hesitações e

²⁴ É um método de construção que faz parte da história brasileira, principalmente na região Norte e Nordeste, que consiste no uso do barro e da madeira para criar moradias.

²⁵ Nos anos anteriores ao início da pesquisa, não foi possível dar início às gravações, uma vez que as restrições impostas pela pandemia da Covid-19 impossibilitaram o acesso aos moradores da comunidade.

²⁶ Implementado na Universidade Federal da Bahia (UFBA), em 2002, pelo seu criador e coordenador, o Prof. Dr. Dante Lucchesi, o Projeto Vertentes tem formado um expressivo acervo da fala popular do Estado da Bahia. São mais de 150 entrevistas, entre comunidades rurais isoladas, comunidades do recôncavo baiano e comunidades urbanas periféricas. Os estudos desenvolvidos a partir desse acervo, prestam relevantes contribuições para a

intervenções pragmáticas típicas da linguagem falada (Chave de Transcrição do Projeto Vertentes, disponível nos anexos desta pesquisa).

5.2 O ENVELOPE DA VARIAÇÃO

Os fatores linguísticos que atuam na variação e mudança linguística, juntamente com os fatores sociais, compõem o envelope da variação. Este é o conjunto das motivações de uma regra variável. Há fatores que condicionam a escolha de uma ou de outra determinada variante. Nesta seção, apresentaremos a **variável dependente**, que é a realização/não realização fonética do sujeito pronominal, apresentando as variantes que compõem essa variável, além das variáveis linguísticas e extralinguísticas que também afetam a variável em estudo, denominadas de **variáveis independentes** ou **explanatórias**.

5.2.1 A variável dependente

A variável dependente - a realização/não realização pronominal de sujeito - neste estudo é binária, ou seja, pode assumir duas possibilidades: presença do pronome sujeito ou ausência de pronome sujeito foneticamente realizado. As ocorrências que serão analisadas compreendem todas as pessoas pronominais, tanto no singular como no plural, em sentenças declarativas e interrogativas. Como este fenômeno é bastante recorrente na língua, o número de dados gerados é bem expressivo, as ocorrências válidas são as sentenças cujos verbos possuem forma finita. Assim, é necessário estabelecer critérios para melhor compreender o comportamento da variável dependente.

5.2.1.1 Critérios para a exclusão de dados

Algumas ocorrências foram descartadas da amostra, pois estas poderiam alterar o universo de observação e distorcer o resultado das análises.

- a. Sujeitos nulos de frases consideradas expressões cristalizadas não foram considerados:
(01) Ø **Vô** *dizê* que **vô**? (RS02)

b. Ocorrências com verbos em formas não finitas não foram consideradas:

(03) Essas menina faz o quê, rapaz? Só **brincá** e **comê**, somente. (RS02)

c. Ocorrências com verbo implícito, ainda que seja possível recuperá-lo através do contexto, foram descartadas:

(04) Doc.: Você sabe fazer caruru?

Inf.: Eu não! (LC01)

d. Sujeito nulos em verbos que funcionam como marcadores discursivos:

(05) ... Esse tempo assim nesse coisa assim eu não lembro não porque às vezes nem contava assim coisa pra mim, **sabe**? (RS02)

e. Construções que possuem um sujeito nulo com valor dêitico:

(06) e todo mundo vivendo e isso aqui crescendo, aí... **ficou dum**, isso aqui grande e quase tempo o povo querendo perder tudo isso aqui... (SS03)

f. Sujeitos pronominais *a gente* e *você(s)* com referência [-definida/+genérico]:

(07) Fazer a casa, cavar, botar enchimento e depois bota a vara, aí vai **cava** o barro, **bota** água ta... Aí **pisa** e **tapa** pra ficar ligado. (RM01)

5.2.1.2 As variáveis linguísticas

Como há muitos fatores que podem alterar a frequência de sujeitos nulos e plenos, apresentaremos uma descrição dos principais fatores internos e externos que podem afetar o fenômeno investigado.

5.2.1.2.1 Pessoa do sujeito

Os estudos realizados por Duarte (1993, 1995) demonstraram, na amostra sincrônica analisada, que havia um movimento em direção ao preenchimento do sujeito, liderada pela primeira e segunda pessoa pronominal. Segundo a autora, a terceira pessoa apresenta um contexto de resistência à implantação da mudança em função de sua relação de correferência com um SN anterior, e a resistência ao uso do pronome ele/a com antecedentes com o traço semântico [- animado]. Ao analisar a realização do sujeito pronominal no português afro-

brasileiro, Lucchesi (2009, p. 177) encontrou resultados semelhantes aos de Duarte (1995, p. 48), apontando que o sujeito pronominal nulo é mais elevado na terceira pessoa, seguida pela primeira pessoa, sendo a segunda pessoa o contexto menos favorável ao sujeito nulo.

Pretende-se verificar, através dessa variável linguística, a frequência de supressão dos sujeitos de primeira e terceira pessoa e a frequência de retenção pronominal para a segunda pessoa, a qual se apresenta como o contexto mais favorecedor para a mudança conforme os estudos anteriores. As formas pronominais que integram esta variável são:

a) Primeira pessoa do singular (eu/ Ø)

- (08) Não, não **eu** não mato e... de junto **eu** não encosto não, que **eu** não *aguento* ver.
Num Ø guento ver por que é a merma coisa de um ser humano.

b) Segunda pessoa do singular (tu/você/ Ø)

- (09) Aí **tu** deixa aí que eu encho de noite.
 (10) Eu não sei se **você** deve conhecer lá um colega meu também que vinha pescar demais aqui.

c) Terceira pessoa do singular (ele/ela/ Ø)

- (11) Mas depois **ela** pegou um *problema* aí **ela** não parou de trabalhar.
 (12) Não, só tenho duas cabeça, uma vaca e uma bezerra que Ø pariu agora.

d) Primeira pessoa do plural (a gente/ Ø)

- (13) **A gente** passava necessidade aí o rio...

e) Primeira pessoa do plural (nós/ Ø)

- (14) **nós** nunca mudou para canto nenhum, sempre **nós** *tamo* convivendo aqui.
 (15) Tirei se **nós** morasse *ni* outro canto e não Ø pagasse aluguel.

f) Segunda pessoa do plural (vocês/ Ø)

- (14) **vocês** não vai me tirar a força, viu?

g) Terceira pessoa do plural (eles/elas/ Ø)

(15) **Eles** começou tirando cerca e $\emptyset$ disse que vinha o trator ia passar pelas casas não sei que... $\emptyset$ *Iludiro* muita gente pra poder ir daqui pra casa minha vida.

5.2.1.2.2 Desinência gramatical do verbal

Segundo Chomsky (1981), um paradigma mais simplificado tende a favorecer a realização do sujeito pronominal. Duarte (1995) constatou que o enfraquecimento do paradigma verbal do PB, desencadeado pela entrada de *você* e *a gente* no sistema pronominal, gerou uma reorganização estrutural que teve como consequência o aumento da frequência de sujeitos realizados, uma vez que o sincretismo das formas verbais não permite que a identificação do pronome sujeito seja feita através da morfologia verbal.

Os dados de Lucchesi (2009, p. 178) revelaram que o contexto mais favorecedor ao sujeito pronominal nulo é a presença do morfema *-mos* da primeira pessoa do plural, que é o morfema número-pessoal mais saliente do paradigma flexional. O princípio da saliência fônica²⁷, proposto por Naro e Lemle (1976), já é consagrado no campo das análises variacionistas no Brasil. Os morfemas da primeira pessoa e o *-m* final ficaram num nível intermediário, enquanto a ausência de qualquer morfema aparece na literatura como o contexto mais resistente ao sujeito nulo. As desinências que fizeram parte do estudo em questão estão descritas abaixo:

As desinências verbais que fizeram parte da análise foram as seguintes:

a. Morfema zero (- pessoa / - plural)

(16) Naquele tempo ela ganhava dinheiro, muito muito muito dinheiro, mas **ela** pegava o dinheiro e **ela** gastava.

b. Marcas de primeira pessoa do singular (+ pessoa / - plural)

(17) **Eu** tenho medo dela cair.

(18) Também só **fui** no médico nesse dia.

²⁷ Segundo Scherre (1989, p.301), o princípio da saliência fônica “consiste em estabelecer que as formas mais salientes, e por isso mais perceptíveis, são mais prováveis de serem marcadas do que as menos salientes”.

c. Marcas de primeira pessoa do plural (+ pessoa / + plural)

(19) A parte triste daqui é que **nós** não podemos nem fazer uma entrevista.

d. Marcas de terceira pessoa do plural (- pessoa / + plural)

(20) **Elas** [filhas] crescerem na *page*, obedecê,

5.2.1.2.3 Caracterização semântica do sujeito

A caracterização semântica do sujeito é apontada como uma das variáveis mais relevantes para o comportamento desse fenômeno. Segundo Cyrino, Duarte e Kato (2000), os sujeitos com o traço semântico [+animado], representados pelas primeiras e segundas pessoas, exibem as maiores frequências de realização fonética, ao passo que os sujeitos de terceira pessoa se localizariam numa escala mais baixa no nível de referencialidade, sendo foneticamente realizados com menor frequência.

Duarte (2019, p. 107) constata que é o traço [-animado] que explica a lentidão na implantação da mudança nos sujeitos de terceira pessoa. Os resultados encontrados por Lucchesi (2009) em quatro comunidades rurais isoladas na Bahia também confirmam os resultados obtidos por Duarte (1995), pois o traço semântico [-animado] favoreceu de forma considerável a não realização do sujeito de terceira pessoa, com um Peso Relativo (PR) de 0.73. Desse modo, como observou Duarte (1995, p. 77) o traço semântico [- animado] tem se constituído em um poderoso fator para o uso de categorias vazias no português brasileiro, tanto no caso dos objetos diretos anafóricos (Omena, 1978; Duarte, 1989; Cyrino, 1993), quanto no caso do incremento das relativas cortadoras (Tarallo, 1983).

Assim, ao analisar o traço semântico do referente do sujeito pronominal nas sentenças finitas dos dados da comunidade Baixa da Linha, buscaremos verificar se o traço [-animado] também se configura como um contexto de resistência à implantação da mudança. Alguns exemplos encontrados nos dados que ilustram esses níveis de referencialidade são exibidos a seguir:

a. [+ animado/ + humano]

(21) Que **ele** [meu pai] morreu, me Ø deixou criado.

b. [+ animado / – humano]

(22) Ah esse peixe nós compra aí na na na rua, né, nos comércio. Agora da onde Ø vem quem sabe?

c. [- animado/ - humano]

(23) (...) levou uns quinze dia, **ele**[o dente] largou dentro da boca.

5.2.1.2.4 Tipo de frase

Duarte (1995) constatou que as orações interrogativas favorecem a não realização do pronome sujeito. Diante dessa constatação, pretendemos verificar se além das interrogativas, outros tipos de frases também influenciam na frequência de realização fonética do sujeito pronominal da comunidade afro-brasileira em questão:

a. Afirmativa

(24) Lá *mermo* **a gente** pescava.

b. Negativa

(25) Mas eu não comia não que **eu** não gosto de jiló.

c. Interrogativa

(26) Pra que **eu** quero terra?

d. Resposta a pergunta direta do tipo yes/no question

(26) Doc.: Mas você gosta de comer?

Inf.: Gosto.

e. Resposta a pergunta direta do tipo yes/no question com sim/não

(27) DOC: O trem... você acha que é perigoso?

INF.: Não, não, Ø só é [o trem] perigoso se você tiver na linha, na hora que ele passar.

f. resposta a uma pergunta QU- (Wh- *Questions*)

(28) DOC: O que você mais gosta daqui da comunidade Baixa da Linha?

INF: **Eu** gosto aqui é do tempo da plantação.

g. Resposta epistêmica

(29) DOC: Mas se te oferecesse uma casa em outro lugar?

INF: **Eu** nem sei, né? Depende.

5.2.1.2.5 Paralelismo formal

Outra variável importante que afeta o comportamento linguístico dos indivíduos foi formalizada por Scherre e Naro (1993) por meio do conceito de paralelismo formal. Para os pesquisadores, o falante tende a manter o mesmo padrão de marcação em orações sequenciadas com o mesmo sujeito sintático (Scherre; Naro, 1993, p. 8-11). Esta tendência de manutenção de um mesmo padrão de marcação no encadeamento da fala sugere que a presença de um sujeito nulo na oração anterior favorece sobremaneira a não realização do sujeito pronominal na oração seguinte (Lucchesi, 2009, p. 180). Lucchesi obteve em sua análise um PR de 0.74 para as ocorrências em que o sujeito pronominal nulo era antecedido por sujeito pronominal apagado, uma vez que o sujeito não realizado na oração anterior favoreceu a manutenção do sujeito nulo na oração subsequente. A não realização do sujeito pronominal também é favorecida quando o antecedente é um SN.

a. Sintagma nominal

(30) **Meu pai** que vivia aqui, mas ele perdeu uma perna.

b. Pronome

(31) Eu **eu** acho que eu tô com problema nas vista.

c. Nulo

(32) **Ø** nasci aqui, eu sou nascida aqui.

d. Primeira menção (ocorrências que não estão ligadas a uma ocorrência

anterior)

(33) **Eu** gosto daqui.

5.2.1.2.6 Localização em relação ao antecedente e as condições de referencialidade

O estudo de Duarte (1993) sobre o PE focalizou a relação entre a função sintática da oração na qual o sujeito se encontra e a frequência de realização do sujeito pronominal. Os dados encontrados pela autora apontaram que, nas orações relativas, os sujeitos foneticamente realizados superaram a frequência dos sujeitos nulos (70% e 61%, respectivamente). Já as orações sem encadeamento sintático favoreceram o uso do sujeito nulo. Esses resultados demonstram a importância de considerar essas variáveis no grupo de fatores deste estudo. O padrão sentencial é outro fenômeno que será analisado paralelamente, assim como a estrutura do Sintagma Complementizador²⁸ (CP).

As variantes que integram essa variável são as seguintes:

a. Sem encadeamento sintático

(34) **Eu** tenho cinco filho, tenho dois homem e três mulher.

b. Primeira oração assindética

(35) Tinha tempo que eu comprava comida, **Ø** botava dentro de casa pra ela comer porque o dinheiro dela sumia.

c. Segunda oração assindética

(36) A gente aqui planta, a gente colhe daqui mesmo, **a gente** veve da plantação.

d. Terceira oração assindética em diante

(37) A gente gosta de morar aqui, a gente cria galinha, cria porco, **cria** gado.

e. Primeira coordenada sindética

(38) Aí a gente fazia aquelas lata de beiju e **Ø** levava.

²⁸ Do inglês *Complementizer Phrase*.

f. Segunda coordenada sindética

(39) Ela pegava e gastava na festa bebendo e Ø ficava dentro de casa com fome.

g. Terceira coordenada sindética em diante

(40) Ele saía e ficava por lá, e eu ficava em casa porque Ø não podia sair mesmo.

h. Oração principal posposta à subordinada em que está o antecedente

(41) Aí quando eu comecei trabalhar **eu** fiquei mais alegre, eu era muito triste.

i. Oração adverbial com antecedente na oração principal

(42) Eu mesmo criei meus filho *quasemente* com o Rio Capivari e tripa de galinha porque **eu** não tinha recurso.

j. Oração completiva com antecedente na oração principal

(43) Aí ela veio aqui e disse que Ø ia ajudar esses povo daqui

k. Oração relativa com antecedente na oração principal

Os vizinho tem uns que Ø percuram umas, uma zuadinha, um pobremazinho...

5.2.1.2.7 Presença de constituinte à esquerda do verbo na oração

Esta variável se refere à presença ou ausência de constituintes na estrutura sintática da oração de forma a verificar se há relevância no comportamento dos sujeitos pronominais em relação aos tipos de constituintes, quando presentes.

a. Oração simples

(44) **A gente** depende de transporte.

b. Com conjunção coordenativa

(45) Outras coisa **eu** não tive muito pá dar, não, mas educação **eu** tenho que dar, né?

c. Com conjunção subordinativa

(46) Esse ano eu tô *pensano* se Ø vou prantar ou não.

d. Com elemento adverbial antes do sujeito

(47) Aqui **eu** não tenho *oto* lugar *p'eu* ir.

e. Com SN antes do sujeito

(48) É Dona Vera **ela** é uma pessoa nascida e criada na comunidade.

(49) Eu... todo ano **eu** pranto amendoim.

f. Com pronome relativo/interrogativo antes do sujeito

(50) Aqui a comida que **nós** mais come bem é feijão.

5.2.1.3 *As variáveis sociais*

Os estudos sociolinguísticos concebem a dimensão social como intrínseca à variação. A língua não deve ser analisada fora das questões que influenciam o seu funcionamento. Os informantes, nesta pesquisa, foram estratificados em categorias sociais específicas (Labov, 2008 [1972]) tradicionalmente utilizadas no âmbito das pesquisas sociolinguísticas: sexo/gênero, faixa etária e escolaridade.

5.2.1.3.1 Sexo

Embora os estudos sociolinguísticos realizados entre as décadas de 1960 e 1970 (Labov, 1966 [2008], 1972, entre outros) apontassem como tendência ao comportamento linguístico feminino o uso de variáveis linguísticas mais conservadoras e de maior prestígio social, os papéis de sociais de homens e mulheres tiveram mudanças significativas comparando-se o período em questão com a atual organização desses papéis na sociedade. Freitag (2015) defende que “se a sociedade muda, as explicações do modelo teórico-metodológico deveriam, também, mudar” (p. 12). Segundo ela, a visão de que as mulheres preferem as formas linguísticas menos estigmatizadas em função de sua condição materna não seja mais válida hoje, diante dos novos papéis desempenhados pelas mulheres na sociedade. Desta forma, uma análise que envolva a variável sexo/gênero deve considerar a estrutura social da comunidade em questão.

Buscaremos testar, segundo a hipótese mais tradicional, se existe, por parte das mulheres, uma maior tendência de uso da variante conservadora - o sujeito não realizado foneticamente. No entanto, como este fenômeno variável não apresenta variantes socialmente estigmatizadas, espera-se que haja neutralização dos resultados em relação à variável sexo/gênero.

5.2.1.3.2 Faixa etária

A faixa etária é uma variável social que permite observar a dinâmica da mudança de determinado fenômeno em uma comunidade. Freitag (2011) afirma que idade, juntamente com classe e sexo, é uma das três supercategorias sociais nas sociedades modernas. A autora argumenta também sobre a percepção da influência da idade em processos de variação e mudança em usos linguísticos que taxamos como “fora de moda”, gírias desatualizadas; em outras palavras, percebemos de forma intuitiva que mesmo que os anos tenham passado, continuamos a manter determinados traços linguísticos em nosso repertório, ainda que, para a comunidade, possam ter caído em desuso (Freitag, 2011, p. 46).

Nesta pesquisa, os informantes foram estratificados em três faixas etárias: Faixa 1: 25 a 35 anos; Faixa 2: 45 a 55 anos; Faixa 3: a partir de 65 anos. A hipótese mais frequente acerca desta variável tradicional no âmbito dos estudos sociolinguísticos é que esta reflete diferentes estágios da mudança linguística. Como as mudanças dentro do escopo da Sociolinguística Variacionista podem ser apreendidas ao longo do curso de implementação - o tempo aparente - que se baseia nas diferenças linguísticas apresentadas entre gerações, podendo, portanto, espelhar mudanças diacrônicas a partir do recorte sincrônico.

O estudo de Almeida (2005, p. 1999) comprova tal afirmação, pois ao comparar seus resultados com outros trabalhos, principalmente de Duarte, verificou-se que esse fenômeno não apresenta o mesmo comportamento em relação a outras variantes já descritas. A distribuição dos sujeitos nulos por faixa etária, por exemplo, evidenciou uma estabilização do fenômeno, pois há um equilíbrio nos percentuais de cvs usadas pelos mais velhos e pelos mais jovens, o que sugere o uso de uma mesma gramática.

5.2.1.3.3 Escolaridade

Geralmente a variável social Escolaridade, no âmbito das pesquisas sociolinguísticas, tem sido associada à resistência às tendências de mudança. Segundo Mollica e Braga (2003, p. 53), “a escola atua como preservadora de formas de prestígio”, isso repercute nos gostos, normas e padrões de dizer e escrever dos indivíduos que nela frequentam. Entretanto, nem todos os fenômenos linguísticos incidem sob o controle normatizador da escola. Enquanto alguns usos são punidos pela escola outros usos são coniventes, considerados “como formas redundantes, do tipo *há anos atrás*” (Mollica e Braga, 2003, p. 53).

Nesta pesquisa, os participantes foram estratificados em analfabetos (pessoas com nenhuma escolaridade) e semianalfabetos (pessoas que tiveram acesso à escolarização por tempo máximo de 4 anos). No entanto, há pouca diferença em relação à escolarização desses grupos, o que também interfere pouco no comportamento linguístico desses informantes. No caso específico do fenômeno desta pesquisa, a realização pronominal do sujeito, a interferência da variável escolaridade é bem menos sensível, uma vez que este fenômeno não sofre as sanções impostas pela escola. Apesar disso, assumimos a possibilidade de que os informantes que tiveram algum contato, mesmo que ínfimo, com a escolarização formal, apresentem tendência a utilizar a variante mais conservadora – o pronome nulo.

5.3 Processamento quantitativo dos dados

A Sociolinguística Variacionista utiliza ferramentas para obter resultados confiáveis em relação a fenômenos em variação. O instrumental quantitativo é, portanto, a base de um modelo de teoria linguística que procura explicar os padrões quantitativos de uso das possibilidades linguísticas (Guy; Zilles, 2007). Nesta pesquisa, a ferramenta estatística escolhida foi o pacote de programas Varbrul (Sankoff, 1988; Pintzuk, 1988), especificamente a versão mais recente, o Goldvarb X (Sankoff, Tagliamonte, Smith, 2005), que apresenta uma metodologia de análise baseada em um método multivariado. Segundo Guy e Zilles (2007, p. 33-34), o método multivariado busca “separar, quantificar e testar a significância dos efeitos de fatores contextuais em uma variável linguística”. Assim, este método de análise é o mais adequado para esta pesquisa, pois, será possível analisar, de modo simultâneo, influência e significância entre as variáveis independentes (Guy, Zilles, 2007).

6 ANÁLISE VARIACIONISTA DA REALIZAÇÃO DO SUJEITO PRONOMINAL NA COMUNIDADE BAIXA DA LINHA-BA

O fenômeno do sujeito nulo é assim um dos mais complexos e difíceis de ser apreendido na sua totalidade, e um dos lugares emblemáticos das contradições linguísticas (...) (Galves, 2019, p. 6).

Nesta seção, serão apresentados os resultados quantitativos dos procedimentos de análise a que foram submetidos os dados do *corpus* no estudo em questão. Além dos percentuais encontrados, apresentaremos uma comparação com os resultados identificados em outras pesquisas de modo a confrontar nossas hipóteses com as hipóteses já levantadas pela literatura, visando compreender o estado atual do fenômeno nesta comunidade e sua similaridade ou dissimilaridade em relação a outras. Sendo assim, busca-se contribuir para o desenvolvimento de estudos descritivos sobre o português brasileiro, especificamente para a variedade do português afro-brasileiro.

Seguindo a metodologia traçada na seção 5, contabilizamos 909 dados em contextos que permitem a variação entre sujeito nulo e sujeito realizado em forma pronominal. Inicialmente, para a análise da variável dependente, controlamos 14 (catorze) grupos de fatores (variáveis independentes) sendo: **variáveis linguísticas**: (1) pessoa do sujeito; (2) desinência gramatical do verbo; (3) caracterização semântica do sujeito; (4) tipo de oração; (5) predicação; (6) menção anterior do sujeito; (7) localização em relação ao antecedente; (8) estrutura sintática da oração; (9) duplo sujeito; e (10) presença de constituinte à esquerda do verbo; e **variáveis sociais**: (11) sexo; (12) faixa etária; (13) escolaridade; e (14) estada fora da comunidade. A partir da rodada dos dados no *software* GoldvarbX, o programa nos apresentou os grupos de fatores mais significativos para o preenchimento da posição de sujeito, ou seja, para a aplicação da regra variável, como é possível observar na tabela 6:

Tabela 6- Grupos de fatores selecionados como significantes para a ocorrência do sujeito pronominal segundo o programa²⁹

Variáveis selecionadas	Fatores que compõem a variável	P / T	(%)	p.r.
Tipo de oração	Declarativa (afirmativa)	413/640	64	0.539
	Negativa	91/137	66	0.530
	Interrogativa	6/10	60	0.385
	Resposta à pergunta direta quando for sim/não	18/55	32	0.176
	Pergunta com sim ou não	4/11	36	0.192
	Resposta em contexto <i>Wh-questions</i>	23/40	57	0.433
	Resposta epistêmica	12/16	75	0.490
Menção anterior do sujeito	Sintagma nominal	41/87	47	0.356
	Pronome	209/328	63	0.559
	Nulo	61/180	34	0.241
	Primeira menção (ocorrências que não estão ligadas a uma ocorrência anterior)	256/314	81	0.639
Localização em relação ao antecedente	Sem encadeamento sintático	375/477	78	0.660
	Outro período	83/151	55	0.533
	1ª oração assindética	18/84	21	0.141
	2ª oração assindética	2/26	7	0.082
	3ª oração assindética em diante	2/11	18	0.218
	1ª coordenada sindética	32/71	45	0.291
	2ª coordenada sindética	10/17	59	0.478
	3ª coordenada sindética em diante	1/1 Nocaute	100	-
	Principal posposta à subordinada em que está o antecedente	13/20	65	0.501
	Adverbial (antecedente na principal)	17/26	65	0.286
	Completiva (antecedente na principal)	3/5	60	0.219
	Relativa (antecedente na principal)	12/21	57	0.345
Estrutura sintática da oração	Oração simples	339/593	57	0.401
	Oração com conjunção coordenativa	57/104	55	0.524
	Oração com conjunção subordinativa	83/109	76	0.688
	Oração com elemento adverbial antes do sujeito	29/31	93	0.835
	Oração com SN antes do sujeito	31/32	97	0.931
	Oração com pronome relativo/interrogativo antes do sujeito	28/40	70	0.552

Fonte: Elaborado pela autora (2025)

Era esperado que algumas variáveis linguísticas e sociais como a pessoa do sujeito, a desinência gramatical do verbo, a caracterização semântica do sujeito e a faixa etária fossem selecionadas pelo programa, no entanto, não foi o que ocorreu. Um possível problema que atuou

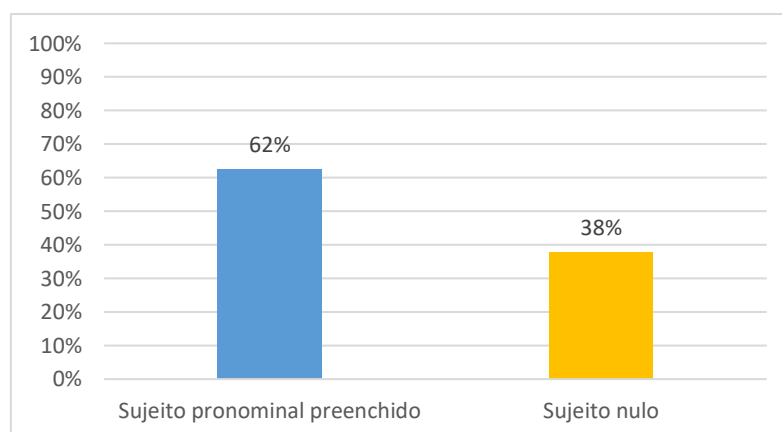
²⁹ Os resultados apresentados na tabela 6 foram inseridos os contextos com orações coordenadas. No entanto, a análise detalhada de cada fator nas subseções posteriores deste estudo, foram baseadas nos dados sem o contexto de orações coordenadas.

nesses grupos de fatores foi a multicolinearidade, uma vez que esses grupos são semelhantes. Desse modo, como alguns grupos estão associados efetivamente a fatores específicos de outros grupos como a primeira menção e a primeira pessoa do discurso, pode ser que um fator acabou impulsionando o outro devido à semelhança entre eles, levando os demais grupos a perderem força estatística. Portanto, focaremos mais na descrição dos resultados encontrados do que necessariamente na análise de pesos relativos.

6.1 A variável dependente

Apresentaremos, a seguir, a análise dos resultados encontrados por meio de gráficos e tabelas dos resultados gerados pelo programa Goldvarb X. Conforme o gráfico apresentado abaixo, é notável a preferência entre os moradores da comunidade Baixa da Linha para o preenchimento da posição de sujeito, de modo que o índice de sujeitos nulos é reduzido:

Gráfico 2- Sujeito pronominal expresso e sujeito nulo

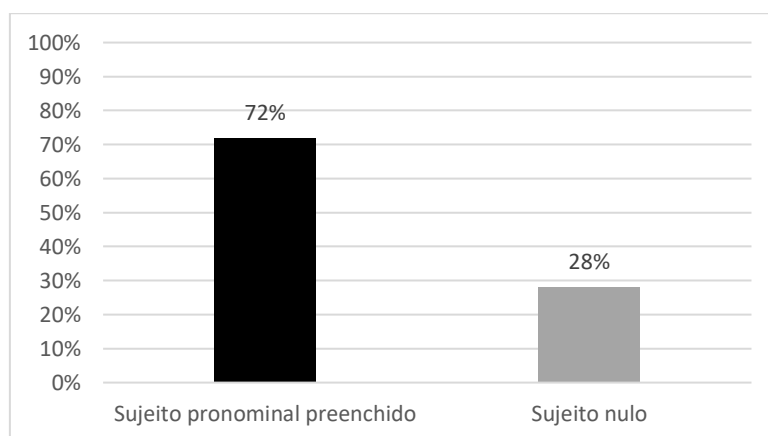


Fonte: Elaborado pela autora (2025)

Os resultados fornecidos pelo programa estatístico apontaram que das 909 ocorrências encontradas, 62% (correspondendo a 567 ocorrências) foram de sujeitos foneticamente realizados (preenchidos), ao passo que 38% dos dados (equivalente a 342 ocorrências) representaram sujeitos não realizados foneticamente (nulos). No entanto, neste resultado foram considerados os contextos de orações coordenadas com sujeitos correferentes, contexto não computado por Duarte (1995, 2018) a partir do segundo membro da coordenação na variedade urbana do Rio de Janeiro, pois, segundo ela, esta não é uma propriedade relacionada ao PSN,

já que, segundo a autora, em línguas [- Sujeito Nulo] é possível o não preenchimento de um sujeito em tais contextos. Sendo assim, realizamos uma nova rodada excluindo as coordenadas não iniciais, visando estabelecer uma comparação com os resultados encontrados por Duarte. No gráfico abaixo, é possível observar os resultados encontrados excluindo-se as ocorrências com orações coordenadas não iniciais.

Gráfico 3: Sujeito pronominal expresso e sujeito nulo sem orações coordenadas



Fonte: Elaborado pela autora (2025)

Excluindo este contexto, o número total de ocorrências foi reduzido para 697. Da análise do gráfico, podemos perceber uma diminuição de 10% no total de sujeitos foneticamente não realizados em comparação aos resultados apresentados no gráfico 2, o que confirma que o contexto de dupla ou tripla coordenação atua como fator bastante relevante para o favorecimento da não expressão do sujeito, uma vez em que basta a expressão do sujeito no primeiro termo da coordenação. O percentual de sujeitos preenchidos, com tal exclusão, foi de 62% para 72%, evidenciando que as primeiras coordenadas tendem a apresentar maior percentual de sujeitos foneticamente realizados enquanto as segundas e terceiras coordenadas apresentam frequência menor de sujeitos expressos; tivemos, portanto, um reequilíbrio dos resultados em linha ao proposto pela literatura, com o aumento percentual dos expressos e diminuição dos nulos. Nos exemplos 01 e 02 é possível perceber orações coordenadas com sujeitos correferentes conforme encontrados nos dados desta pesquisa:

(01) **Eu** vou nos dois, **Ø** vou no *Candombré*, **Ø** vou na igreja (...). (RM01)

(02) (...) **ele** [irmão] saía, **Ø** ia pra qualquer lugar e **Ø** ficava por lá. (SS03)

Apresentaremos, a seguir, os resultados do estudo sem o contexto das coordenadas, ainda que realizemos algumas comparações com e sem esse padrão sentencial. Observamos, para estabelecer tal comparativo, os resultados de Duarte (1995, 2012, 2019), referentes à frequência do sujeito nulo no PB urbano culto e sobre o estudo diacrônico do sujeito em peças de teatro; de Lucchesi (2004, 2009); de Almeida (2005) e de Silva (2023), sobre o português afro-brasileiro. A tabela 7 apresenta a frequência de sujeitos nulos e preenchidos de referência definida encontradas nesses estudos.

Tabela 7- Percentual de sujeitos expressos *versus* sujeito nulo

	Sujeito expresso %	Sujeito nulo %	Sujeitos nulos/Total
Duarte (1995)	71	29	415/1.424
Lucchesi (2004, 2009)	73	27	1.255/4.599
Almeida (2005)	69	31	1.797/5.802
Silva (2023)	65	35	369/1.056

Fonte: Adaptado de Duarte (1995), Lucchesi (2004, 2009), Almeida (2005) e Silva (2023).

Os dados apresentados na tabela 7 evidenciam uma tendência relativa no PB no que diz respeito à alternância entre sujeito expresso e sujeito nulo. Observa-se que, em todos os estudos analisados (Duarte, 1995; Lucchesi, 2004, 2009; Almeida, 2005; Silva, 2023), há predominância do sujeito expresso, com percentuais que variam entre 65% e 73%. O sujeito nulo, por sua vez, apresenta taxas entre 27% e 35%, ainda que minoritária, essa ocorrência é significativa.

Os resultados da literatura apresentados na tabela 7 revelam percentuais muito semelhantes para a variável dependente, na direção de um percentual muito maior do preenchimento do sujeito em relação ao sujeito não foneticamente realizado. É importante ressaltar, no entanto, que no estudo de Lucchesi (2004, 2009), os contextos com orações coordenadas foram computados, ao passo que nas demais pesquisas foram desconsiderados os contextos com coordenadas não iniciais, o que pode explicar o maior percentual de sujeitos expressos, 73%, e o menor índice de sujeitos nulos, 27%; essa, no entanto, é uma diferença marginal, já que todos os estudos apresentam resultados semelhantes: preenchimento acima de 65% e nulos abaixo de 35%. Nota-se, ademais, que o estudo mais recente, Silva (2023) apresenta ligeiro aumento na taxa de sujeito nulo (35%), o que pode indicar reconfigurações no sistema pronominal.

Alguns pesquisadores têm caracterizado o português brasileiro como uma língua de sujeito nulo parcial (Duarte, 1995; Kato, 1999; Galves, 2001), uma vez que, embora preserve a possibilidade de omissão do sujeito pronominal em determinados contextos, apresenta aumento significativo da realização fonética do pronome, sobretudo em comparação com o PE.

A frequência de sujeitos foneticamente realizados e de sujeitos nulos encontrados entre falantes da comunidade Baixa da Linha, como vemos, são congruentes com os resultados de outras pesquisas, como Duarte (2018 [1995], 2019), Lucchesi (2004, 2009), Almeida (2005) e Silva (2003). Nossa pesquisa, assim, revela o favorecimento do sujeito pleno em relação ao sujeito não preenchido. Seguindo a variedade urbana culta do PB, o PAF tem preenchido a posição de sujeito pronominal e apresentado índices inclusive menores de sujeitos nulos, apontando que talvez haja ali um processo de mudança mais adiantado. A partir desses resultados, não podemos afirmar categoricamente que o PB tem se mostrado exatamente como uma língua de sujeito nulo parcial (Duarte; Marins, 2021), já que os resultados mostram que essa língua vem se encaminhando num processo de mudança em direção ao preenchimento, devido à ausência de uma distribuição equânime nos dados, o que têm evidenciado que o PB vem preenchendo cada vez mais a posição de sujeito em suas variedades.

A comparação entre os dados da comunidade Baixa da Linha e o comportamento do sujeito nulo em línguas tipologicamente classificadas como de sujeito nulo parcial, como o finlandês, revela alguns pontos de convergência. Em ambos os casos, observa-se que a omissão do sujeito não ocorre de forma categórica, coexistindo com o uso de sujeitos pronominais explícitos, conforme Holmberg (2005):

- i. (Minä) puhun englantia
- ii. (Eu) falo inglês
- iii. (Sinä) puhut englantia
- iv. (Tu/você) falas inglês
- v. (Me) puhumme englantia
- vi. (Nós) falamos inglês
- vii. (Te) puhutte englantia
- (Vocês) falam inglês
- viii. *(Hän) puhuu englantia
- Ele/ela fala inglês
- ix. *(He) puhuvat englantia
- Eles falam inglês

Nota-se que o finlandês permite a omissão do pronome sujeito quando a flexão verbal identifica claramente a pessoa, sobretudo na primeira e segunda pessoas, no entanto, o sujeito não pode ser livremente omitido livremente na terceira pessoa conforme exemplos (viii e ix), pois a sentença torna-se agramatical. Diferentemente do finlandês, em que a omissão do sujeito é favorecida pela riqueza da morfologia flexional verbal, sobretudo nos contextos de primeira e segunda pessoa, no PB esse fenômeno se desenvolve em um contexto de erosão morfológica e reorganização do sistema pronominal, o que tem levado ao aumento do preenchimento da posição de sujeito. As razões para o alto preenchimento da posição de sujeito no PB se relacionam a fatores históricos associados à erosão da morfologia flexional causada pelas implicações do contato linguístico.

No entanto, o estudo de Magalhães (2006) revelou que as crianças produzem sentenças gramaticais sem sujeito expreso, mesmo recebendo no *input* tanto estruturas com sujeito nulo quanto com sujeito preenchido. A autora identificou a ocorrência recorrente de sentenças produzidas por crianças sem sujeito foneticamente realizado, comprovada por análise quantitativa de fala espontânea infantil, como nos exemplos abaixo coletados pela autora em situações cotidianas, como brincadeiras, conversas e atividades rotineiras.

Ø quer água.

Ø caiu.

Ø pegou.

O sujeito não aparece foneticamente, mas é recuperável pelo contexto discursivo, geralmente referindo-se à própria criança ou a um participante já mencionado. A análise sustentada pela autora é que a frequência de sujeitos não expressos é maior na fala das crianças do que na fala adulta em determinados contextos, correspondendo a cerca de 60% das ocorrências, enquanto os sujeitos expressos correspondem a aproximadamente 40% dos dados. Torna-se, portanto, necessário o desenvolvimento de mais estudos sobre sentenças produzidas por crianças em contexto de aquisição de modo a analisar os fatores que favorecem ou inibem a realização do sujeito pronominal.

O percentual acentuado de sujeitos pronominais realizados na comunidade Baixa da Linha, 73%, revela que a gramática dessa comunidade não é a de uma língua [+ Sujeito Nulo], ainda que não prototipicamente a de uma língua [- Sujeito Nulo], como o inglês, que apresenta sujeito expletivo e índices de preenchimento próximo aos 95%. Esse comportamento nos leva

a inferir que, no *input* disponível às crianças, há um número reduzido de dados em que a posição de sujeito esteja sem uma representação fonética, o que pode fazer com que esses adquirintes passem a fixar o parâmetro com base nesses dados que, de forma consistente, apresenta o sujeito preenchido, percebendo o preenchimento do sujeito como uma obrigatoriedade.

6.2 As variáveis linguísticas

Apresentaremos, a seguir, alguns resultados encontrados para: pessoa do sujeito; desinência gramatical do verbo; caracterização semântica do sujeito e a presença de constituinte à esquerda do verbo na oração. Apesar dessas variáveis não terem apresentado força estatística para o fenômeno nesta comunidade, ou seja, não tenham sido selecionadas pelo programa, elas são relevantes para ampliar a discussão sobre o comportamento desse fenômeno no PB e estão entre aquelas que, na literatura, têm se mostrado pertinentes para as discussões teóricas sobre o tema. Após a apresentação desses resultados, exibiremos os fatores variáveis encontrados que exercem força estatística para a realização do sujeito pronominal, os fatores selecionados pelo programa, na comunidade rural afro-brasileira Baixa da Linha, em Cruz das Almas-BA, que são: tipo de oração; menção anterior do sujeito; localização em relação ao antecedente; e a estrutura sintática da oração.

6.2.1 Pessoa do sujeito

Uma das variáveis que integra o grupo de fatores deste estudo é a pessoa gramatical do sujeito, pois esta interfere na expressão do sujeito devido ao papel do paradigma pronominal para esse fenômeno, como visto anteriormente. Assim como proposto na escala de referencialidade de Cyrino, Duarte e Kato (2000), os dados apresentados na tabela indicam que o sujeito pronominal nulo é mais elevado para a terceira pessoa, seguido pela primeira pessoa; sendo a segunda pessoa do discurso o contexto menos favorável ao sujeito nulo, conforme se observa na tabela 8.

Tabela 8: Sujeitos expressos x nulos com relação às pessoas gramaticais

Pessoa gramatical	expressos	nulos	total	% nulos
1ª pessoa	337	115	452	25%
2ª pessoa	8	1	9	11%
3ª pessoa	100	71	171	41%
Total	445	187	632	29%

Fonte: Elaborado pela autora (2025)

A primeira pessoa apresenta 452 ocorrências, sendo 337 sujeitos expressos e 115 nulos, o que corresponde a 25% de sujeitos nulos. Esse resultado indica uma tendência predominante ao preenchimento do sujeito, embora a omissão também ocorra com frequência significativa. A possibilidade de sujeito nulo nessa pessoa pode ser explicada pelo fato de o referente ser facilmente recuperável no contexto discursivo, o que permite a omissão sem prejuízo para a interpretação.

Na segunda pessoa foram registradas apenas 9 ocorrências, com 8 sujeitos expressos e 1 nulo, correspondendo a 11% de sujeitos nulos, o menor percentual entre as pessoas gramaticais. Esse resultado sugere que a segunda pessoa constitui um contexto pouco favorável à omissão do sujeito, possivelmente porque o pronome pode desempenhar funções pragmáticas importantes na interação, como marcar diretamente o interlocutor ou enfatizar o destinatário do enunciado.

A terceira pessoa apresenta 171 ocorrências, das quais 100 são expressas e 71 nulas, totalizando 41% de sujeitos nulos, o maior percentual entre as três pessoas gramaticais. Esse resultado indica que a terceira pessoa é o contexto mais favorável à realização de sujeitos nulos. Isso pode ser explicado pelo fato de que o referente da terceira pessoa não participa diretamente da situação de fala e, portanto, sua identificação depende mais do contexto discursivo, o que favorece a omissão quando o referente já foi previamente mencionado.

Em síntese, a pessoa gramatical que exibiu um índice mais elevado de sujeito nulo foi a terceira pessoa, enquanto a primeira, e, principalmente a segunda pessoa, a frequência de categorias vazias na posição de sujeito foi bem reduzida no *corpus*. Como não se observa o mecanismo da correferência na segunda pessoa, a mesma também não dispõe de morfemas flexionais específicos no PB, portanto, manifesta-se como o contexto mais favorecedor para a retenção pronominal (Lucchesi, 2009, p. 177). Isso evidencia que, no quadro do processo de mudança, a terceira pessoa se mostra como um contexto de resistência, enquanto nas outras duas, em especial a segunda, a mudança segue em vias de implementação. Vejamos abaixo os resultados encontrados para cada pessoa do sujeito separadamente.

Tabela 9: Sujeitos expressos relacionados às pessoas gramaticais separadamente

Pessoa gramatical	expresso	total	%
1ª pessoa do singular	317	430	74
2ª pessoa do singular	8	9	89
3ª pessoa do singular	90	144	62
1ª pessoa do plural	75	87	88
2ª pessoa do plural distintiva (vocês)	2	2	100
			Nocaute
3ª pessoa do plural	10	27	38

Fonte: Elaborado pela autora (2025)

De modo geral, os resultados indicam que, embora o preenchimento do sujeito seja predominante no *corpus*, a omissão ocorre de forma diferenciada entre as pessoas gramaticais, sendo mais frequente na terceira pessoa. Esse comportamento reforça a influência de fatores discursivo-referenciais na realização do sujeito pronominal no português brasileiro.

As reflexões de Petter (2008) sobre a história sociolinguística do PB também ajudam a contextualizar esses resultados, ao destacar que a formação das variedades afro-brasileiras envolve processos complexos de contato linguístico, nos quais determinadas estruturas gramaticais podem ter sido reorganizadas ou simplificadas ao longo do tempo. Nesse sentido, a maior frequência de sujeitos expressos pode ser interpretada como parte de um processo mais amplo de redução das propriedades prototípicas de línguas de sujeito nulo pleno.

Não houve nenhuma ocorrência de sujeito nulo nos dados selecionados para a segunda pessoa do plural (vós/vocês) e apenas uma ocorrência com a segunda pessoa do singular com pronome sujeito referencial nulo:

(03) Inf.: Ø Sabe quem é que tinha pena da *rente*? (VR03)

Doc.: Não.

Inf.: Era Lula!

As outras 8 (oito), do total de 9 (nove) ocorrências, para a segunda pessoa do singular, foram preenchidas pelo pronome sujeito não-distintivo (tu). Não houve, portanto, nenhuma ocorrência de sujeito nulo com o pronome canônico (tu). Resultados semelhantes foram encontrados por Almeida (2005, p. 104) nas comunidades rurais dos municípios baianos de

Bananal/Barra dos Negros, Mato Grosso e Matinha, apresentando apenas 1/11 ocorrências de sujeito nulo para a segunda pessoa do singular e 2/29 para a segunda pessoa indireta do plural.

A segunda pessoa, representada pelos pronomes *tu*, *vós*, *você(s)*, nos quais a concordância se dá de forma mais frequente com a marca de terceira pessoa nessa variedade do PB, apresentou os maiores percentuais de preenchimento no estudo de Duarte (1995), Almeida (2005) e Lucchesi (2009). Esta baixa frequência da não expressão do sujeito pode ser explicada porque normalmente nesses contextos “não se observa o mecanismo da correferência e não se dispõe de morfemas flexionais específicos no PB” (Lucchesi, 2009, p. 177). Assim, de acordo com Duarte (1995, p. 49), o alto preenchimento da posição de sujeito na segunda pessoa é quase categórico, devido a substituição pelos pronomes de tratamento *você(s)*, *senhor(es)*, incorporados rapidamente ao sistema pronominal e partilhando marcação morfológica com a terceira pessoa. Entretanto, nos estudos desenvolvidos por Lucchesi (1994, 2001a, 2002a, 2002b, 2006a), a frequência de uso do sujeito nulo nessas comunidades afro-brasileiras já havia sido restringida devido a erosão de sua morfologia flexional, ainda mais adiantada que nas variedades outras do português, comportamento esse determinado pela intensidade diferenciada dos processos de TLI em sua formação.

Os resultados da tabela 9 indicam, ainda, 26% de ocorrências de sujeito nulo para a primeira pessoa do singular, a única que possui desinência distintiva, o que revela, apesar da morfologia, uma frequência cada vez maior de preenchimento, 74% para primeira do singular e 89% para o plural. Inserindo o contexto das orações coordenadas, não há grandes alterações no quantitativo geral dos resultados, já que encontramos um índice de 66% de preenchimento para a primeira pessoa do singular e de 88% para a primeira pessoa do plural, ilustrados nos exemplos abaixo:

(04) Aí Ø tive meus filho tudo naquela casa lá ali. (SS03)

(05) **Eu** trabalho em casa de farinha também. (RM01)

(06) **Você** conhece uma *mulé* que é vereadora, saudar, saudar[...] Botô as menina tudo pá *trabaiá*. Ia botá as minina tudo pá *trabaiá*. Só quis o voto! (VR03)

(07) “**Tu** vai ficar aqui na terra!”

O reduzido número de dados para a segunda pessoa, tendo em vista o grande quantitativo para a primeira, pode ser explicado pela própria estrutura da entrevista, na qual os participantes narraram suas vivências na comunidade sob uma perspectiva muito pessoal, o que

diminui as possibilidades de direção do diálogo para o entrevistador, exigindo o uso de uma forma de referência a ele. Constatamos, assim, que o contexto de segunda pessoa, ainda que tenha um preenchimento grande, precisa ser complementado por mais estudos nessas comunidades que envolvam métodos de recolha de dados para além da entrevista sociolinguística padrão.

Em relação à terceira pessoa, nossos resultados indicam o que outros pesquisadores já tiveram como principais achados sobre o sujeito nulo, sendo este contexto o mais propício para a não realização, com 62% de preenchimento para a terceira pessoa do singular e 38% para a terceira pessoa do plural, como nos exemplos abaixo:

(08) Não, não. Só Ø é [o trem] perigoso se você tiver na linha, na hora que ele passar que aí que é perigoso, mas fora isso, não Ø é não.

(09) É uma semana pra Ø tá nascendo. É! Com cinco dia **ela** [mandioca] já tá toda nascidinha.

(10) Todos **eles** [pais] nasceu aqui, meu pai e minha mãe, **eu** só não fiz *conhecê* minha mãe.

(11) Aqui **eu** tenho cinco [irmãos]. Era sete... sete, morreu dois. Aí só ficou cinco! Ø Mora tudo aqui! Ø Nasceu tudo aqui! (RS02)

Os resultados obtidos para o fator pessoa do discurso podem ser interpretados à luz da hierarquia de referencialidade proposta por Cyrino, Duarte e Kato (2000). Nessa perspectiva, quanto menor o grau de referencialidade do referente, maior a tendência de realização nula do pronome. Assim, a maior frequência de sujeitos nulos na terceira pessoa confirma essa previsão, uma vez que essa pessoa apresenta menor grau de referencialidade em relação às pessoas do discurso. Já a primeira pessoa ocupa posição intermediária, enquanto a segunda pessoa, diretamente associada ao interlocutor e frequentemente marcada por funções pragmáticas de interação, apresenta menor favorecimento à omissão do sujeito.

Em suma, a distribuição dos sujeitos pronominais observada neste estudo revela um padrão que dialoga diretamente com resultados já descritos na literatura sobre o português brasileiro. No *corpus* analisado, verifica-se predominância de sujeitos expressos (71%) em

relação aos nulos (29%), com comportamento diferenciado entre as pessoas gramaticais: a terceira pessoa apresenta o maior percentual de sujeitos nulos (41%), seguida pela primeira pessoa (25%), enquanto a segunda pessoa registra apenas 11%, configurando-se como o contexto menos favorável à omissão. Esse padrão converge, em linhas gerais, com os resultados apresentados por Almeida (2005) em seu estudo sobre comunidades do interior da Bahia, no qual a autora também identifica que a realização do sujeito pronominal é sensível à pessoa gramatical, sendo a terceira pessoa o contexto mais propício à omissão, sobretudo quando o referente já se encontra estabelecido no discurso. Ao mesmo tempo, os resultados aqui obtidos corroboram as análises de Duarte (1995, 2003), que apontam para uma mudança em curso no português brasileiro caracterizada pela redução progressiva do sujeito nulo e pela expansão do sujeito pronominal expreso, fenômeno associado à erosão do paradigma flexional verbal e ao enfraquecimento das propriedades típicas de línguas de sujeito nulo pleno.

Nesse sentido, a predominância de sujeitos expressos observada na comunidade quilombola Baixa da Linha pode ser interpretada como evidência de alinhamento com essa tendência mais ampla do português brasileiro contemporâneo. Por outro lado, a persistência de índices relativamente elevados de sujeitos nulos, especialmente na terceira pessoa, também pode ser compreendida à luz dos estudos de Lucchesi (2009) sobre a formação histórica das variedades populares do PB, particularmente aquelas associadas a comunidades rurais afrodescendentes. Segundo o autor, processos de TLI e intenso contato linguístico contribuíram para a configuração de sistemas gramaticais que apresentam padrões próprios de variação morfossintática.

Assim, os resultados encontrados neste estudo parecem refletir a intersecção entre dois processos complementares: de um lado, a tendência geral do português brasileiro ao aumento do preenchimento do sujeito, conforme descrito por Duarte; de outro, a permanência de estratégias de omissão condicionadas por fatores históricos e discursivo-referenciais, especialmente na terceira pessoa, padrão também observado em variedades rurais analisadas por Almeida (2005) e interpretado, no quadro histórico mais amplo, pelas propostas sociolinguísticas de Lucchesi sobre a constituição do português popular brasileiro.

6.2.2 Desinência gramatical do verbo

Dentre as quatro formas de morfologia verbal, apenas uma mostrou-se favorecedora do sujeito nulo, (-*m*), a qual se aplica à segunda pessoa do plural e à terceira pessoa do plural. As

desinências verbais associadas à primeira pessoa do singular e do plural foram as que mais favoreceram o sujeito pronominal realizado nos dados encontrados, como podemos ver na tabela abaixo. Isso indica o estágio avançado deste fenômeno na estrutura linguística desta variedade do PB.

Tabela 10- O sujeito expresso e a desinência gramatical do verbo

Desinência gramatical do verbo	Ocorrências	Total	%
- pessoa / - plural (-0 / -ou)	233	334	70
+ pessoa / - plural (-o / -ei)	266	360	74
+ pessoa / + plural (-mos)	4	4	100
- pessoa / + plural (-m)	1	3	Nocaute 33

Fonte: Elaborado pela autora (2025)

Os dados da tabela 10 indicam que o sujeito expresso é predominante em todos os contextos de desinência verbal, embora em graus diferentes. Do ponto de vista funcional, esperava-se que desinências mais distintivas de pessoa e número favorecessem a omissão do sujeito, uma vez que a flexão verbal poderia recuperar o referente, nota-se, no entanto, pela análise da tabela, que com a desinência número-pessoal [+ pessoa / + plural], não houve nenhuma ocorrência de sujeito pronominal nulo. Este resultado, portanto, não confirma a hipótese inicialmente apresentada, pois esperava-se que morfemas mais salientes, como o (-mos), aplicado à segunda pessoa do plural, pudessem dispensar a necessidade de representação do sujeito pronominal devido à possibilidade de retomada via flexão verbal. Todavia, como essas quatro ocorrências foram encontradas na fala de uma mesma informante, analfabeta, faixa etária 3, é possível que seja apenas um comportamento linguístico individual e não uma tendência geral da comunidade em relação a esta desinência verbal. Sendo assim, um maior número de dados pode de fato fornecer uma visão mais abrangente sobre esta questão, ressaltando ainda mais a necessidade de novos estudos sobre essas comunidades.

Os resultados obtidos, apesar disso, confirmam o que se apresenta na literatura em relação à primeira pessoa do singular, pois a desinência [+ pessoa / - plural] é o contexto em que há mais preenchimento do sujeito, com apenas 26% de sujeitos nulos, apesar de sua morfologia. Isso demonstra novamente que mesmo em verbos com desinência latente, o sujeito expresso é foneticamente realizado.

A interpretação do alto índice de sujeitos pronominais expressos em relação à desinência número-pessoal do verbo pode ser articulada com as propostas teóricas de Lucchesi (2001, 2009, 2015) sobre o contato linguístico e a TLI na formação do português popular brasileiro. Nesse quadro, a redução da transparência das marcas morfológicas de pessoa e número no paradigma verbal pode enfraquecer o papel da flexão verbal como mecanismo exclusivo de identificação do sujeito. Como consequência, os falantes tendem a recorrer com maior frequência a estratégias explícitas de referência, entre elas o uso do sujeito pronominal expresso, que garante maior clareza na identificação do referente no discurso. Assim, mesmo em contextos em que a desinência verbal poderia teoricamente recuperar a pessoa do sujeito, observa-se uma alta taxa de preenchimento pronominal, como verificado nos dados da comunidade quilombola Baixa da Linha.

Desse modo, o elevado percentual de sujeitos expressos associado às diferentes desinências número-pessoais do verbo pode ser interpretado como reflexo de um sistema linguístico em que a flexão verbal perdeu parte de sua força distintiva, favorecendo o uso de elementos pronominais para a explicitação do referente. Esse comportamento está em consonância com as análises de Lucchesi, segundo as quais as variedades populares do português brasileiro apresentam traços estruturais que resultam de processos históricos de contato linguístico e de transmissão irregular da língua, os quais contribuíram para a reconfiguração do sistema morfossintático, incluindo os mecanismos de marcação da referência pronominal.

Nesse sentido, os dados analisados sugerem que, na variedade investigada, a presença de marcas flexionais de pessoa e número no verbo não constitui um fator determinante para a omissão do sujeito, uma vez que o preenchimento pronominal permanece elevado em praticamente todos os contextos morfológicos. Esse padrão reforça a hipótese de que, nessa variedade do português brasileiro, estratégias discursivas de explicitação referencial desempenham papel central na organização da estrutura do sujeito, em consonância com o quadro histórico-sociolinguístico proposto por Lucchesi.

Conforme argumenta Lucchesi (2009, 2015), as variedades populares do português brasileiro, especialmente aquelas associadas a comunidades rurais afrodescendentes, foram formadas em contextos de contato linguístico e de TLI. Esses processos históricos contribuíram para a reorganização de diferentes aspectos da gramática, incluindo fenômenos morfossintáticos relacionados à marcação de pessoa e número. Nesse contexto, o uso mais frequente de sujeitos expressos pode funcionar como estratégia de explicitação referencial,

compensando eventuais ambiguidades decorrentes da redução ou neutralização de marcas morfológicas.

Assim, os dados da comunidade quilombola Baixa da Linha sugerem que a presença ou ausência de marcas flexionais de pessoa e número no verbo exerce influência limitada na realização do sujeito, uma vez que o preenchimento pronominal permanece elevado mesmo em contextos morfológicamente distintivos.

6.2.3 Caracterização semântica do sujeito

Segundo Barbosa, Duarte, Kato (2005), a animacidade é um dos fatores de grande relevância para a compreensão da variação do sujeito nulo. Enquanto o traço [+ animado] favorece a manutenção do pronome, o traço semântico [- animado] é um fator que inibe a presença do pronome. Para pesquisadores como Duarte, 1995, 2019a; Almeida, 2005; Almeida; Carneiro, 2009; Lucchesi, 2009c, o traço [-humano] atua como um fator que retarda a mudança, sendo considerando um contexto de resistência para a mudança paramétrica no Brasil. Assim, observemos os seguintes exemplos, que ilustram a chamada hierarquia de referencialidade para o preenchimento do sujeito proposta por Cyrino, Duarte e Kato (2000).

[+ animado / - humano]

(12) Ø É[cobra] um bicho que... quando **ela** vê a gente **ela** ‘tá correndo da gente aí a gente corre dela aí é dois medroso.

(13) Aí quando ‘tava nascendo [plantação] não Ø deixa[a galinha] um... uma *raíge* que Ø cisca e Ø come, Ø *tora* todo quando ‘tá nascendo.

(14) **Eu** mato logo! Quem quer conta?
Não. O povo não tem o dizer que **ela** [a cobra] vem e busca a pessoa na cama?

[+ animado / + humano]

(15) **Eu** gosto aqui é do tempo da plantação, porque **nós** planta tudo.

(16) Hoje tem energia, tem água, antes **a gente** pegava água da fonte.

(17) Morreu um homem que se Ø **jogou** lá [na linha]. Ø **Tava** lá *dethado*, aí **ele** morreu.

(18) Meu **pai** que vivia aqui, mas **ele** perdeu uma perna.

[- animado / - humano]

(19) Já ‘Ø tava rachando as madeira já, Ø tava podre e aí Ø caiu.

(20) Só tinha um rio ali que Ø matava a fome da rente.

(21) **Ele** [macarrão] fica solto, **ele** fica com gosto do tempero mesmo, do alho, do corante, tudo que tiver que nós botá.

Em nossos dados, os resultados se mostraram em linha com o encontrado pela literatura, como podemos ver na tabela abaixo:

Tabela 11– Sujeito expresso e caracterização semântica do sujeito

Caracterização semântica do sujeito	Ocorrências	Total	%
[+ animado / - humano]	2	12	16
[+ animado / + humano]	491	669	73
[- animado / - humano]	7	16	44

Fonte: Elaborado pela autora (2025)

A análise da variável caracterização semântica do sujeito revela que o contexto que mais favorece a realização do sujeito pronominal expresso é o conjunto [+animado / +humano], com 73% de preenchimento. Esse resultado está em consonância com a hierarquia de referencialidade proposta por Cyrino, Duarte e Kato (2000), segundo a qual referentes humanos ocupam posições mais altas na escala referencial e tendem a favorecer a realização pronominal explícita. Nessa perspectiva, participantes humanos do discurso são mais salientes do ponto de vista pragmático e discursivo, o que aumenta a probabilidade de que sejam retomados por meio de pronomes sujeitos realizados foneticamente.

Diferente do esperado, o segundo contexto de maior favorecimento foi o de sujeitos [-animado / - humano]; isso, no entanto, pode ser explicado pelo baixo número de dados, apenas 16, o que pode distorcer os índices percentuais. O contexto que mais favoreceu os sujeitos nulos, diferente do proposto pela literatura, foi o contexto [+ animado / - humano]; no entanto, o baixo número de dados, apenas 12, pode estar impedindo que os índices percentuais apresentem um retrato verdadeiro. Dessa forma, embora alguns resultados apresentem divergências em relação às tendências descritas na literatura, possivelmente em razão do número reduzido de

ocorrências em determinados contextos, os dados analisados confirmam a relevância do traço de animacidade na variação do sujeito pronominal.

Ao analisar dados do português popular rural baiano, Almeida (2005) observa que o traço semântico de animacidade, especialmente associado ao traço [+humano], desempenha papel central na organização do sistema pronominal e na realização do sujeito no português brasileiro. Em seu estudo, sujeitos humanos tendem a apresentar maior saliência discursiva e maior frequência de realização explícita, enquanto sujeitos inanimados mostram comportamento distinto, frequentemente associados a estruturas em que a referência é recuperável pelo contexto discursivo.

Os dados do presente *corpus* confirmam essa tendência: a predominância de sujeitos com traço [+animado / +humano] (73%) indica que a realização do sujeito se concentra majoritariamente em referentes humanos, reforçando a hipótese de que a animacidade atua como fator estruturante na distribuição do sujeito no PB. Essa convergência empírica reforça a interpretação de que o sistema do PB privilegia referentes humanos na estrutura sintática, o que pode ser relacionado tanto à reorganização do PSN, discutida por Cyrino, Duarte e Kato, quanto aos efeitos sócio-históricos associados à formação do PB, como propõe Lucchesi.

Assim, os dados aqui analisados corroboram a ideia de que fatores semântico-discursivos, como a animacidade, são decisivos para compreender a dinâmica entre sujeito nulo e sujeito expresso no PB, particularmente em variedades populares e historicamente marcadas por processos de contato linguístico. Assim, a análise desse fator linguístico evidenciou a necessidade de ampliar as investigações empíricas em comunidades afro-brasileiras, de modo a compreender com maior precisão como fatores estruturais, discursivos e históricos interagem na configuração do sistema pronominal do PB nessas variedades.

6.2.4 O tipo de oração

A análise dos dados desta pesquisa foi realizada a partir de entrevistas realizadas com os moradores da comunidade. Como a pesquisa procurou obter e analisar o vernáculo da comunidade, as perguntas versavam sobre a experiência de vida e permitiam que os participantes se sentissem confiantes para relatar suas vivências. Desse modo, o total de sentenças declarativas (como em 22 e 23) foi muito mais expressivo comparado aos demais tipos, seguido pelas orações negativas (26 e 27) e em contextos de resposta à pergunta direta do tipo sim/não. Esses resultados encontram-se apresentados na tabela a seguir:

Tabela 12- Sujeito expesso em relação ao tipo de oração

Tipo de oração	Ocorrências	Total	%	p.r
Declarativa	359	475	75	0.54
Negativa	81	102	80	0.57
Interrogativa	6	9	67	0.51
Resposta epistêmica	12	16	75	0.44
<i>Wh-questions</i>	22	34	65	0.40
Resposta a pergunta direta sim/não	17	51	33	0.17
Pergunta com sim/não	3	10	30	0.13

Fonte: Elaborado pela autora (2025)

A análise da realização do sujeito expesso em função do tipo de oração na comunidade Baixa da Linha revela diferenças importantes associadas ao grau de autonomia informacional da estrutura discursiva. Conforme os dados apresentados, as orações negativas (80%) e declarativas (75%) exibem as maiores taxas de realização do sujeito expesso, seguidas das respostas epistêmicas (75%) e das orações interrogativas (67%). Em contrapartida, observa-se uma redução significativa da expressão do sujeito em contextos como *wh-questions* (65%), respostas a perguntas diretas sim/não (33%) e perguntas polares com sim/não (30%), sendo estes últimos os contextos de menor favorecimento, como indicam também os pesos relativos mais baixos (0.17 e 0.13).

Nos contextos de perguntas polares e respostas a perguntas diretas, observa-se forte redução da realização do sujeito expesso. Isso ocorre porque tais estruturas estão altamente dependentes do contexto interacional imediato, no qual a referência ao sujeito já se encontra disponível no discurso precedente. Assim, a omissão do sujeito não compromete a recuperação referencial, o que favorece a ocorrência de sujeito nulo. Além disso, a maior taxa de sujeito expesso em respostas epistêmicas (75%), ainda que com peso relativo intermediário (0.44), pode indicar que esse tipo de construção frequentemente envolve marcação de posicionamento do falante ou avaliação modal do conteúdo proposicional, o que favorece a explicitação do sujeito como estratégia de reforço da responsabilidade enunciativa.

Diferentemente dos dados obtidos por Duarte (1995), as frases interrogativas não constituíram um contexto preferencial para o apagamento do sujeito. Em seus dados, a autora encontrou 77% de sujeitos nulos em contraposição aos 67% encontrados nas orações declarativas. Entrevistas baseadas em diálogos não favorecem a incidência de indagações por parte do entrevistado, uma vez em que se busca ouvir a narrativa de vida dos informantes; logo,

o quantitativo de ocorrências para orações interrogativas é bem reduzido. Apesar disso, é possível notar uma inclinação ao preenchimento do sujeito, como em (24). Conforme a tabela 12, os dados apontam um certo favorecimento do sujeito pronominal nas frases negativas e declarativas, conforme p.r. de 0.57 e 0.54, respectivamente; no caso das sentenças negativas, a partícula de negação localizada na posição pré-verbal exerce um maior favorecimento na expressão do sujeito pronominal, que podemos ver em (27), apresentando 0.57 de p.r.

(22) Não, porque **eu** enjoei mesmo aí **eu** não como.

(23) Toda vez que **nós** partia [...] **ele** chamava pra partir castanha, **ele** me *alevava* na rua pra vendê.

(24) **Você** conhece uma *mulé* que é vereadora? [...]

(25) **ØTenho** tanta terra aqui! Pra que **eu** quero terra?

(26) Não **Øtenho** o que falar, graças a Deus, aqui o lugar é sossegado. (DN01)

(27) Ah... Mas **eu** nem sei muito *cuzinhar*, só **Øsei** *cuzinhar* o arroz, o feijão. (LC01)

6.2.5 Menção anterior do sujeito

Esta variável diz respeito ao paralelismo formal (Scherre e Naro, 1993), que consiste na repetição de escolhas linguísticas tanto no nível discursivo (i. é, intersentencial) quanto no nível sintático (i. é, intrassentencial). Esses autores defendem a tese de que há uma tendência de manutenção pelos falantes de certas formas linguísticas desde que não haja mudança no referente. Baseado nessa tendência, Lucchesi (2009) verificou a atuação desse fator em seu estudo sobre o PAF, e encontrou que o sujeito não realizado favoreceu o sujeito nulo na oração seguinte, uma preservação da estrutura. Em nosso estudo, encontramos os seguintes resultados:

Tabela 13– Sujeito expresso e menção anterior do sujeito

Menção anterior do sujeito	Ocorrências	Total	%	p.r.
Primeira menção (Ocorrências que não estão ligadas a uma ocorrência anterior)	253	308	82	0.67
Pronome	167	217	77	0.51
Nulo	49	111	44	0.18

Sintagma nominal		31		61		51		0.25
------------------	--	----	--	----	--	----	--	------

Fonte: Elaborado pela autora (2025)

A análise da variável menção anterior do sujeito evidencia a forte influência do encadeamento discursivo na realização do sujeito pronominal. Como se observa na Tabela 13, o contexto que mais favorece o sujeito expresso é o de primeira menção ao referente no discurso, com 253 ocorrências de sujeito preenchido em um total de 308 casos, correspondendo a 82% de realização e peso relativo de 0.67. Esse resultado sugere que, quando um referente é introduzido pela primeira vez no discurso, há uma tendência significativa à explicitação do sujeito, o que pode ser interpretado como uma estratégia de estabelecimento e identificação do referente para o interlocutor.

Na sequência, observa-se que o contexto em que o sujeito da oração anterior é um pronome expresso também favorece a manutenção do preenchimento do sujeito, apresentando 77% de realização e peso relativo de 0.51. Esse resultado indica um efeito de persistência estrutural ou paralelismo discursivo, no qual a presença de um pronome em uma oração tende a favorecer a repetição dessa mesma estratégia nas orações subsequentes, contribuindo para a continuidade da estratégia de explicitação referencial.

Por outro lado, os contextos em que a ocorrência anterior apresenta sujeito nulo ou sintagma nominal mostram-se menos favoráveis ao preenchimento pronominal. O contexto de sujeito nulo anterior apresenta apenas 44% de sujeitos expressos e peso relativo de 0.18, configurando-se como o fator que mais favorece a manutenção do sujeito nulo nas orações subsequentes. Já os casos em que a menção anterior ocorre por meio de sintagma nominal apresentam 51% de sujeitos expressos e peso relativo de 0.25, indicando um comportamento intermediário.

De modo geral, esses resultados evidenciam que a realização do sujeito pronominal na comunidade investigada é fortemente condicionada por fatores discursivos e de continuidade referencial, demonstrando que a forma de expressão utilizada em um determinado ponto do discurso tende a influenciar a escolha estrutural nas orações subsequentes. Esse padrão confirma a importância do encadeamento referencial na alternância entre sujeito nulo e sujeito expresso, revelando que o fenômeno não depende apenas de fatores morfosintáticos, mas também da organização informacional do discurso. Detalhamos, a seguir, os resultados por fator:

a. Sintagma nominal

Esta variável se refere aos sujeitos de terceira pessoa a partir da correferência no discurso. A hipótese inicial era a de que a presença de um sintagma nominal antecedente desfavoreceria a realização do sujeito pronominal, uma vez que o referente já estaria explicitamente introduzido no discurso, o que poderia favorecer a retomada por meio de sujeito nulo.

Os resultados obtidos, no entanto, não confirmam plenamente essa hipótese. O fator apresentou 51% de sujeitos expressos, com peso relativo de 0.25, indicando um comportamento próximo da neutralidade, embora com leve tendência ao desfavorecimento do sujeito exposto. Esses dados sugerem que, mesmo quando o referente já foi introduzido por um sintagma nominal, os falantes ainda recorrem com frequência à explicitação do sujeito pronominal, como ilustrado em (28). Esse resultado aponta para o fato de que a presença de um antecedente lexical não elimina necessariamente a necessidade de retomada explícita do referente, possivelmente como estratégia de manutenção da clareza referencial no discurso.

Por outro lado, também se observam casos em que a retomada ocorre por meio de sujeito nulo, como em (29), indicando que a alternância entre formas nulas e expressas permanece disponível nesse contexto. A proximidade entre os percentuais de realização e não realização do sujeito sugere que esse fator exerce influência limitada na escolha da variante, o que pode estar relacionado ao número relativamente reduzido de ocorrências identificadas.

De modo geral, os resultados indicam que a presença de um sintagma nominal antecedente não constitui um fator determinante para o apagamento do sujeito na variedade analisada, reforçando a ideia de que a realização do sujeito pronominal no português brasileiro é fortemente condicionada por fatores discursivos e referenciais. Ainda assim, a confirmação mais robusta dessa tendência depende da ampliação do *corpus*, uma vez que um conjunto mais amplo de dados poderia evidenciar com maior precisão o papel desse fator na alternância entre sujeito nulo e sujeito exposto.

(28) **Minha filha** me acaba {risos}, mas **ela** não sabe o que é o amor, né? O amor é pra quem sabe. (MC03)

(29) Porque quando a **escola** veio pra aqui **Ø**tomou muita terra do povo. (MC03)

b. Pronome

Quando numa sentença, a ocorrência anterior é preenchida por um pronome expreso, há um favorecimento para que haja continuidade dessa representação pronominal na oração seguinte, o que configura o caso mais prototípico do paralelismo formal. Nesses contextos, encontramos 77% das ocorrências de sujeitos preenchidos por pronomes plenos, com o peso relativo de 0.51. Esse resultado sugere que a presença de um pronome em uma oração anterior exerce influência significativa sobre a estrutura da oração subsequente, promovendo a continuidade da realização pronominal. Em outras palavras, a ocorrência pronominal contribui para a manutenção da referência explícita ao sujeito ao longo do encadeamento discursivo.

A sentença em (30) ilustra esse padrão, em que a repetição do pronome eu reforça a continuidade referencial e evidencia o efeito de paralelismo estrutural:

(30) **Eu** trabalho em casa de farinha também. **Eu** torro também. (RA01)

c. Nulo

Nos contextos em que a ocorrência anterior apresenta sujeito nulo, observa-se uma tendência à manutenção dessa mesma estratégia na oração subsequente, configurando um padrão de paralelismo estrutural semelhante ao observado nos casos de pronomes expressos. Esse comportamento indica que a forma de realização do sujeito em um ponto do discurso pode influenciar diretamente as escolhas estruturais nas orações seguintes.

Nos dados analisados, esse fator apresentou 44% de sujeitos expressos e peso relativo de 0.18, o que indica forte desfavorecimento do sujeito expreso e, conseqüentemente, favorecimento da continuidade do sujeito nulo. Em outras palavras, quando o sujeito é omitido em uma oração, há uma tendência de que essa omissão seja mantida nas sentenças subseqüentes, especialmente em contextos de continuidade referencial, nos quais o referente já se encontra claramente identificado no discurso.

O exemplo em (31) ilustra esse padrão, no qual sucessivas ocorrências de sujeito nulo mantêm a referência ao mesmo tópico discursivo, sem necessidade de retomada explícita:

(31) **Eles** começou tirando cerca e **Ødisse** que vinha o trator. **ØIa** passar pelas casas não sei que... **ØIludiro** muita gente. (MC02)

Esse comportamento evidencia que a alternância entre sujeito nulo e sujeito expreso não depende apenas de fatores morfossintáticos, mas também da organização discursiva e da gestão da referência ao longo do enunciado. Assim, quando o referente já está estabelecido e permanece como tópico discursivo, a omissão do sujeito torna-se uma estratégia eficiente de economia linguística e de continuidade informacional.

d. Primeira menção

O fator primeira menção refere-se aos contextos em que o referente do sujeito é introduzido pela primeira vez no discurso, não estando vinculado a uma ocorrência anterior no encadeamento textual. Nesses casos, espera-se que haja maior tendência à realização explícita do sujeito, uma vez que o falante precisa estabelecer de forma clara a referência discursiva para o interlocutor.

Os resultados obtidos confirmam essa expectativa. Esse fator apresentou 82% de sujeitos expressos, com peso relativo de 0.67, configurando-se como o contexto que mais favorece o preenchimento do sujeito pronominal. Esse comportamento pode ser interpretado como uma estratégia discursiva de introdução e identificação do referente, garantindo que o interlocutor reconheça de forma inequívoca o elemento que passa a ocupar a posição de tópico no discurso. O exemplo a seguir ilustra esse padrão:

(32) Aqui, **eu** nasci aqui. (LC01)

Nesse caso, a presença do pronome *eu* desempenha um papel fundamental na introdução explícita do referente, estabelecendo o sujeito que será potencialmente retomado nas sequências discursivas posteriores. Assim, a realização do pronome na primeira menção contribui para a organização da referência no discurso e para a manutenção da clareza comunicativa. De modo geral, esses resultados reforçam a importância dos fatores discursivo-pragmáticos na alternância entre sujeito nulo e sujeito expreso. A necessidade de introduzir um novo referente no discurso tende a favorecer a explicitação do sujeito, enquanto contextos de continuidade referencial

podem favorecer a omissão. Assim, os dados da tabela 13 evidenciam a confirmação da hipótese para esta variável. Em contextos em que o sujeito é mencionado pela primeira vez no turno de fala, há um expressivo favorecimento de uso do sujeito realizado foneticamente. Isso pode ser explicado pelo fato de ser o contexto inicial e haja a necessidade de especificar o sujeito, realizando uma ancoragem da referência ao objeto discursivo.

A forte tendência ao preenchimento do sujeito em contextos de primeira menção observada na comunidade quilombola investigada pode ser interpretada a partir dos estudos sobre o PAF e os processos históricos de contato linguístico e TLI. Nesse cenário, observa-se frequentemente o fortalecimento de estratégias de explicitação referencial, como a maior realização de sujeitos pronominais, especialmente em contextos discursivos nos quais o referente precisa ser claramente estabelecido.

Assim, o elevado índice de sujeitos expressos em contextos de primeira menção, identificado nos dados da comunidade quilombola Baixa da Linha, pode ser interpretado como reflexo de um padrão discursivo que privilegia a clareza na introdução dos referentes, possivelmente relacionado à configuração histórica do português afro-brasileiro e aos processos de variação e mudança que marcaram a formação dessas variedades linguísticas.

6.2.6 Localização em relação ao antecedente

Duarte (1993) relacionou a função sintática da oração e do elemento retomado à frequência de realização do sujeito pronominal. Em nossa pesquisa, este fator foi selecionado como estatisticamente relevante para a análise do fenômeno. Na tabela abaixo, apresentamos os resultados encontrados para a variável localização em relação ao antecedente.

Tabela 14— Sujeito pleno e localização em relação ao antecedente

Localização em relação ao antecedente	Ocorrências	Total	%	Peso relativo
Sem encadeamento sintático	373	475	78	0.55
Outro período	83	151	55	0.42
Principal posposta à subordinada em que está o antecedente	13	20	65	0.39
Relativa	11	20	55	0.28

Adverbial	17	26	65	0.23
Completiva	3	5	60	0.18

Fonte: Elaborado pela autora (2025)

O contexto que mais favorece a realização do sujeito pleno é o de sem encadeamento sintático, com 78% de sujeitos expressos e peso relativo de 0.55, conforme exemplo abaixo:

(33) *Iludiro* muita gente pra poder ir daqui pra “casa minha vida” (...), **eu** que nunca me interessei. (MC, 02)

Esse resultado sugere que, quando não há relação sintática direta entre as orações, isto é, quando a retomada ocorre em um novo segmento discursivo ou em estruturas menos integradas, os falantes tendem a recorrer com maior frequência ao sujeito expresso. Nesses casos, a explicitação pronominal pode funcionar como estratégia de reforço da identificação referencial, evitando ambiguidades decorrentes da ausência de conexão sintática imediata com o antecedente.

O contexto outro período também apresenta frequência relativamente elevada de sujeitos expressos (55%, p.r. 0.42), o que reforça a hipótese de que maior distância discursiva entre antecedente e retomada favorece a explicitação do sujeito.

(34) *Os pessoal da UFRBi* começou tirando cerca e disse que vinha o trator (...). Aí **eles** *vinhero* aqui me iludindo dizendo que lá era um lugar muito bom.

Quando o referente precisa ser recuperado após a interrupção do encadeamento sintático, a realização do pronome contribui para restabelecer de forma clara a referência no discurso.

Nos contextos de principal posposta à subordinada em que está o antecedente e de orações adverbiais, observam-se percentuais semelhantes de realização do sujeito pleno (65%), embora com pesos relativos mais baixos (0.39 e 0.23, respectivamente). Esses resultados sugerem que, apesar da integração sintática entre as orações, a reorganização da ordem informacional pode levar o falante a recorrer ao sujeito expresso como forma de garantir a continuidade referencial.

(35) Quando *elai* conseguiu as telha, aí Ø cobriu a casa, né? (MC02).

(36) Quando *eui* fiz sete ano **eu** fiquei aqui num jeito que eu tive que trabalhar, com sete ano, porque eu não tinha meu pai, nunca vi, não conheci. (SS03)

Por outro lado, os contextos de orações relativas (55%, p.r. 0.28) e, sobretudo, de orações completivas (60%, p.r. 0.18) mostram menor favorecimento estatístico ao sujeito expresso.

(37) Os *vizinhoi...* tem uns que Ø percuram umas, uma zuadinha (...). (RA02)

(38) Veio aí [os pessoal da *FRB*] que era escola *gronomia* começou tirar cerca e Ø disse que ia tirar muita gente daqui (...). (MC02)

Essa tendência pode ser explicada pelo fato de que essas estruturas apresentam forte integração sintática com a oração antecedente, o que facilita a recuperação do referente por meio de mecanismos de coesão estrutural, reduzindo a necessidade de explicitação pronominal.

Em geral, nota-se um acentuado índice de pronomes nulos associados ao fator localização em relação ao antecedente. Em relação ao contexto de relativas nesses dados, tais como foram encontrados por Galves (2019) nas atas escritas por africanos e afrodescendentes da sociedade protetora dos desvalidos, o contexto das orações relativas é praticamente avesso à presença do sujeito nulo na gramática do PB moderno (Galves, 2019, p. 397), o que nos ajuda a compreender a especificidade das variedades do português afro-brasileiro frente a um comportamento geral, ressaltando a importância da realização de mais estudos sobre essas comunidades.

6.2.7 Estrutura sintática da oração

Para a estrutura sintática da oração, outra variável selecionada como estatisticamente relevante, nossos resultados apresentam o seguinte:

Tabela 15– Sujeito realizado e estrutura sintática da oração

Estrutura sintática da oração	Ocorrências	Total	%	Peso relativo
Oração simples	311	463	67	0.40
Com conjunção coordenativa	22	26	85	0.71
Com conjunção subordinativa	81	107	76	0.57
Com elemento adverbial antes do sujeito	28	30	93	0.81
Com SN antes do sujeito	31	32	97	0.93
Com pronome relativo/interrogativo antes do sujeito	27	39	70	0.48

Fonte: Elaborado pela autora (2025)

A variável estrutura sintática da oração permite observar em que medida a organização interna da sentença influencia a realização do sujeito pronominal. Os resultados indicam que determinados arranjos sintáticos favorecem de forma significativa a explicitação do sujeito, sugerindo que a posição de elementos no interior da oração pode afetar diretamente a estratégia de referência adotada pelos falantes.

O contexto que apresentou maior favorecimento para o sujeito expresso foi o de orações com sintagma nominal antes do sujeito, com 97% de realização e peso relativo de 0.93. Esse resultado indica que, quando um sintagma nominal antecede o sujeito, há forte tendência à explicitação pronominal, possivelmente como forma de evitar ambiguidades interpretativas ou de garantir a identificação clara do referente em estruturas nas quais outros constituintes já ocupam a posição inicial da oração.

Em seguida, destacam-se os contextos em que há elemento adverbial antes do sujeito, com 93% de sujeitos expressos e peso relativo de 0.81, bem como as orações com conjunção coordenativa, que apresentam 85% de realização e peso relativo de 0.71. Esses resultados sugerem que a presença de constituintes prévios ao sujeito tende a favorecer sua realização explícita, uma vez que o deslocamento do sujeito da posição inicial da oração pode aumentar a necessidade de marcação referencial mais clara.

As orações com conjunção subordinativa também apresentam favorecimento ao sujeito expresso, com 76% de realização e peso relativo de 0.57, indicando comportamento relativamente próximo da média geral do *corpus*. Já as orações com pronome relativo ou interrogativo antes do sujeito apresentam 70% de realização e peso relativo de 0.48, sugerindo influência mais moderada desse fator. Conforme demonstrado pela literatura, observa-se que a ocorrência de sujeito nulo é bem expressiva em estruturas com pronome relativo ou

interrogativo anteposto ao sujeito. Esse resultado, em linha com Duarte (1995), reforça que existe uma manutenção dos nulos em contextos mais marcados sintaticamente, tendência que também é constatada na comunidade estudada.

Por fim, as orações simples apresentam 67% de sujeitos expressos, com peso relativo de 0.40, configurando-se como o contexto menos favorável à realização do sujeito pronominal entre os fatores analisados. Esse resultado pode ser interpretado à luz do fato de que, em estruturas sintaticamente mais simples e lineares, a identificação do referente tende a ser mais facilmente recuperada pelo contexto discursivo, reduzindo a necessidade de explicitação pronominal.

De modo geral, os resultados indicam que a realização do sujeito pronominal na comunidade investigada é sensível à configuração sintática da oração, especialmente à presença de elementos que antecedem o sujeito. Em estruturas nas quais outros constituintes ocupam posições iniciais, como sintagmas nominais, advérbios ou conectores, observa-se maior tendência à explicitação do sujeito, possivelmente como estratégia de reforço da identificação referencial e de organização da informação na sentença.

6.3 As variáveis sociais

Os estudos linguísticos já realizados demonstraram que não há grande influência das variáveis sociais para a atual configuração do fenômeno do sujeito nulo. De fato, este padrão se manteve na variedade afro-brasileira da comunidade Baixa da Linha, já que o programa não associou nenhuma variável social ao fenômeno do sujeito nulo.

6.3.1 Sexo

Apesar do programa não haver selecionado esta variável, apresentaremos, na tabela 16 o percentual de sujeitos preenchidos distribuídos pelo sexo dos participantes da pesquisa.

Tabela 16– Sujeito foneticamente realizado e sexo dos informantes

Sexo	Ocorrências com coordenadas /		Ocorrências sem coordenadas /	
		%		%
Feminino	323 / 520	62%	282 / 391	72%
Masculino	244 / 389	63%	218 / 306	71%

Fonte: Elaborado pela autora (2025)

A tabela mostra que, em relação à variação do sujeito pronominal nesta comunidade, a variável sexo não é muito relevante. Esse resultado é reforçado pelos índices percentuais, que mostram que tanto homens quanto mulheres na comunidade possuem padrões semelhantes de preferência pelo preenchimento do sujeito.

Nos contextos com coordenadas, as mulheres apresentam 62% de sujeitos realizados (323 ocorrências em 520), enquanto os homens apresentam 63% (244 ocorrências em 389). A diferença de apenas um ponto percentual sugere que o comportamento linguístico dos dois grupos é praticamente idêntico nesse contexto. De modo semelhante, nos contextos sem coordenadas, as mulheres apresentam 72% de sujeitos realizados (282 em 391) e os homens 71% (218 em 306), novamente revelando uma diferença mínima.

Esses resultados indicam que o fator sexo não exerce influência significativa sobre a realização do sujeito na comunidade investigada, sugerindo que o fenômeno se distribui de maneira relativamente homogênea entre os falantes. Em outras palavras, homens e mulheres compartilham padrões semelhantes no que diz respeito à alternância entre sujeito nulo e sujeito expresso.

Do ponto de vista sociolinguístico, esse comportamento pode ser interpretado como um indício de que a realização do sujeito pronominal não constitui uma variável socialmente marcada por gênero nessa comunidade. Em muitas situações de variação linguística, diferenças entre homens e mulheres podem indicar processos de mudança em andamento ou a associação de determinadas variantes a estilos ou identidades sociais específicas. No entanto, a ausência de diferenças relevantes nos dados sugere que a realização do sujeito se configura como um padrão estrutural relativamente estável na variedade analisada, independentemente do sexo do falante. Esse resultado também reforça a hipótese de que a alternância entre sujeito nulo e sujeito expresso na comunidade quilombola Baixa da Linha é condicionada principalmente por

fatores linguísticos e discursivos e não diretamente por fatores sociais como o sexo dos informantes.

6.3.2 Faixa etária

A variável faixa etária, por sua vez, apresentou os seguintes resultados:

Tabela 17- Sujeito realizado e faixa etária dos informantes

Faixa etária	Ocorrências com coordenadas / %		Ocorrências sem coordenadas / %	
Faixa 1	145 / 225	64%	124 / 175	71%
Faixa 2	228 / 398	57%	215 / 323	67%
Faixa 3	194 / 286	68%	161 / 199	81%

Fonte: Elaborado pela autora (2025)

Nos contextos com coordenadas, observa-se que a faixa 3 apresenta o maior índice de sujeitos realizados (68%, com 194 ocorrências em 286), seguida pela faixa 1 (64%, com 145 ocorrências em 225). A faixa 2 apresenta o menor percentual (57%, com 228 ocorrências em 398). Esse padrão indica uma distribuição relativamente próxima entre as faixas etárias, embora com leve redução na frequência de sujeitos realizados entre os falantes intermediários.

Nos contextos sem coordenadas, as diferenças tornam-se mais evidentes. A faixa 3 apresenta 81% de sujeitos realizados (161 ocorrências em 199), índice significativamente superior ao das demais faixas etárias. A faixa 1 apresenta 71% (124 ocorrências em 175), enquanto a faixa 2 apresenta 67% (215 ocorrências em 323). Esses resultados mostram que os falantes mais velhos tendem a realizar o sujeito com maior frequência, especialmente em contextos sem coordenação.

De modo geral, os dados sugerem que os falantes da faixa 3 mantêm uma tendência mais forte à explicitação do sujeito, enquanto os falantes da faixa 2 apresentam índices relativamente menores de realização. A faixa 1 ocupa uma posição intermediária entre esses

dois comportamentos. Esse padrão pode indicar que a realização do sujeito não está necessariamente associada a um processo linear de mudança linguística em curso, mas pode refletir diferenças de estilo discursivo ou de organização da informação entre gerações.

Apesar da análise dos resultados para as faixas etárias apontar uma nítida preferência pelo sujeito pleno, a faixa 3, representada pelas pessoas com mais de 65 anos, é a que mais tem realizado foneticamente o sujeito pronominal.

Esses resultados indicam que a hipótese da TLI pode realmente explicar o comportamento dessas comunidades, já que esses falantes, mais velhos, tiveram mais contato, através de seus pais e antepassados, aos reflexos das línguas de substrato. As línguas bantu apresentam características morfossintáticas que podem contribuir para a compreensão de certos padrões observados em variedades afro-brasileiras do português. Uma dessas características é o papel central dos marcadores pronominais e concordâncias nominais na estrutura da frase, frequentemente expressos por meio de prefixos de concordância associados ao verbo. Em muitas línguas bantu, como o quimbundo e o quicongo, o sujeito é marcado de forma explícita por meio de prefixos pronominais, que funcionam como índices de pessoa e classe nominal no verbo. No quimbundo, por exemplo, encontram-se estruturas como *ngí-zola* (eu amo), em que *ngí-* indica a primeira pessoa do singular, e *u-zola* (ele ama), em que *u-* marca a terceira pessoa (Chatelain, 1888; Nurse; Philippson, 2003).

De modo semelhante, no quicongo observa-se, de acordo com Bentley (1887), *mono ke-dia* (a criança come), em que o marcador verbal concorda com o sujeito nominal. Esse sistema de concordância explícita reforça a presença de marcas referenciais claras na estrutura da oração. Considerando que grande parte das populações africanas trazidas para o Brasil era falante de línguas bantu, alguns autores defendem que padrões de marcação explícita do sujeito podem ter influenciado a organização de variedades populares do português brasileiro em contextos de contato linguístico (Petter, 2009; Lucchesi, 2009, 2015; Baxter, Lucchesi e Ribeiro, 2009). Nesse sentido, os resultados encontrados na comunidade quilombola Baixa da Linha, no Recôncavo baiano, particularmente o alto índice de preenchimento do sujeito pronominal entre os falantes da faixa etária mais elevada, podem ser interpretados à luz desse histórico sociolinguístico. A preferência por sujeitos pronominais explícitos pode refletir, ainda que indiretamente, estratégias de referenciação mais transparentes, semelhantes às observadas em sistemas linguísticos africanos, nas quais a identificação do sujeito tende a ser formalmente marcada na estrutura da oração. Nesse contexto, ao adquirir a língua portuguesa em um contexto no qual foi presente a transmissão irregular, esses indivíduos preservaram traços importados

das línguas bantu que se refletiram em suas gramáticas. Esse comportamento, por sua vez, foi transmitido de geração em geração.

CONSIDERAÇÕES FINAIS

O português brasileiro apresenta perda gradual das propriedades clássicas das línguas de sujeito nulo consistente, mantendo ocorrências condicionadas estruturalmente. O PB tende, portanto, ao preenchimento do sujeito, embora não tenha perdido totalmente a possibilidade de sujeito nulo. Os dados desta pesquisa mostram uma predominância de sujeitos preenchidos (mais de 70%) na fala da comunidade quilombola Baixa da Linha. Esse resultado é compatível com estudos sobre o PB contemporâneo, que apontam uma forte tendência ao preenchimento do sujeito, sem que a língua tenha perdido totalmente a possibilidade de sujeito nulo. O PB apresenta perda gradual das propriedades clássicas das línguas de sujeito nulo consistente, mantendo ocorrências condicionadas estruturalmente.

Os resultados desta pesquisa revelam uma predominância de sujeitos preenchidos (aproximadamente 70%) na fala da comunidade quilombola analisada. Esse padrão se aproxima das tendências observadas no PB contemporâneo, que apresenta aumento da realização do sujeito pronominal. Contudo, a interpretação desses dados deve considerar o contexto histórico e sociolinguístico das comunidades afrodescendentes, cuja formação está diretamente associada ao contato linguístico e ao processo de Transmissão Linguística Irregular, conforme discutido por Lucchesi e Baxter (2009). Nesse sentido, a variedade analisada pode ser compreendida como parte do *continuum* das variedades afro-brasileiras do português.

Esta pesquisa teve como principal objetivo analisar o comportamento, em relação ao preenchimento do sujeito pronominal, de falantes da comunidade afro-brasileira Baixa da Linha, localizada na cidade de Cruz das Almas-BA. Essa comunidade, falante de uma variedade linguística estigmatizada, sofre os reflexos diretos, no plano linguístico, da marginalização social em que vivem os descendentes de africanos escravizados no Brasil. Investigar o funcionamento linguístico desta comunidade vai para além da identificação dos condicionamentos linguísticos e sociais que interferem nas realizações, mas em promover a visibilidade para uma variedade, uma história e uma cultura representativa de uma parcela da população que teve direitos negados e que continua lutando pela sobrevivência nos quilombos contemporâneos.

Procurou-se, por esses motivos, nos primeiros capítulos deste estudo, apresentar os aspectos sócio-históricos do Recôncavo Baiano, bem como a relevância social e histórica da

comunidade Baixa da Linha, ainda não estudada, no rol dos estudos linguísticos. Esta, como muitas comunidades que se constituem como quilombos contemporâneos e remanescentes de quilombos tradicionais, estão ainda em fase de reconhecimento da sua identidade e direitos. Estes espaços são, portanto, grandes celeiros de ideias, falares e histórias que podem ajudar a compreender processos do passado colonial brasileiro e apontar tendências futuras.

No capítulo 3, buscamos apresentar uma discussão sobre o fenômeno de investigação desta pesquisa, dando ênfase ao processo histórico de Transmissão Linguística Irregular como fator fundante da polarização linguística no Brasil (Lucchesi, 2009, 2019). Apresentamos, em seguida, os aspectos metodológicos da pesquisa, a partir da gravação de 12 entrevistas sociolinguísticas, das quais 10 integraram a amostra, entre informantes estratificados em três faixas etárias, sexo, escolaridade e nascimento na comunidade.

A perspectiva adotada para a análise desta pesquisa foi a Sociolinguística Variacionista, mas também nos debruçamos sobre a Teoria de Princípios e Parâmetros, através da Sociolinguística Paramétrica, de modo a buscar compreender a mudança sintática como um reflexo da remarcação de parâmetros durante o processo de aquisição. Das variantes analisadas, constatamos uma nítida preferência pelo sujeito pronominal preenchido, o qual é favorecido principalmente pelos contextos linguísticos: tipo de oração, menção anterior do sujeito, localização em relação ao antecedente e estrutura sintática da oração. Mesmo com o programa não selecionando nenhum fator social para a atuação do fenômeno do sujeito nulo, a faixa etária apresentou resultados passíveis de análise que precisam ser mais bem investigados, em especial por uma análise das narrativas dos falantes mais velhos, uma vez que é a faixa etária 3, a mais elevada, apresentou os índices mais elevados de preenchimento da posição de sujeito, com 81% de preenchimento e 68% nos contextos com as orações coordenadas.

Esses dados nos levam a confirmar a hipótese inicial adotada nesta pesquisa de que o contato do português, sobretudo com as línguas africanas, afetou a estruturação das variedades linguísticas do PB, especialmente a variedade popular, pois, o processo de reorganização dos pronomes sujeitos nesta variedade já teria sido restringida devido à erosão da morfologia flexional determinada pelo processo de TLI de tipo leve em sua formação. Isso é evidenciado a partir dos altos índices de preenchimento do sujeito entre os falantes mais antigos da comunidade. Além disso, é importante considerar que os falantes mais velhos da comunidade tendem a apresentar menor mobilidade espacial, permanecendo com maior frequência no interior da própria comunidade e mantendo menor contato com centros urbanos. Tal característica sociogeográfica contribui para que esses informantes sofram menor influência de

variedades urbanas do português. Esse aspecto se relaciona com a classificação da comunidade Baixa da Linha na tipologia proposta por Santana, Araujo e Freitag (2018), segundo a qual comunidades quilombolas podem ser caracterizadas como [$\pm$ isolada isoladas. No caso da Baixa da Linha, o relativo grau de isolamento social dos informantes da pesquisa pode contribuir para a preservação de determinados padrões linguísticos observados entre os falantes mais antigos.

Os fatores selecionados pelo programa como estatisticamente relevantes para a variação do sujeito pronominal expresso e para o sujeito não realizado foram: tipo de oração, principalmente nas orações afirmativas e negativas, com o peso relativo 0.54 e 0.57, respectivamente; a menção anterior do sujeito, cuja primeira menção e a presença de um pronome numa oração antecedente favorecem a realização fonética do sujeito, o que comprova a hipótese da atuação do paralelismo sintático nesta comunidade; a localização em relação ao antecedente e a estrutura sintática da oração.

Enquanto nas variedades cultas do PB a frequência do pronome sujeito vem aumentando expressivamente (Tarallo, 1993; Duarte, 1993, 1995, 2018, 2019), tal processo já se encontra em estado avançado nas variedades populares, como é o caso da comunidade da Linha. A baixa relevância da morfologia verbal para a frequência de realização ou não realização do sujeito foi uma das considerações deste estudo, pois tínhamos por hipótese que a flexão verbal seria fundamental para recuperar o sujeito, força que não percebemos nesta comunidade. Entretanto, acreditamos que a realização do sujeito e o paradigma flexional dos verbos tenham sua importância no momento inicial da variação/mudança em direção à preferência pelos sujeitos pronominais realizados. Como o processo já se mostra aparentemente avançado nessa comunidade, outros fatores passaram a atuar juntamente, de forma que o pronome expresso é a variável preferencial dos membros da comunidade rural afro-brasileira Baixa da Linha, de modo que se mostra a forma preferida para as pessoas do sujeito independente da forma verbal utilizada.

O contato linguístico, fator que pode ser visto indiretamente, ganha maior relevância quando consideramos aqueles falantes mais velhos. Estando acima de 65 anos no ano de 2025, esses informantes nasceram antes dos anos 1960, de modo que podem ter tido contato indireto, por seus pais e avós, a antepassados que chegaram a ser escravizados, ou mesmo das últimas levas de escravizados trazidos para o Brasil. Suas histórias e experiências, assim, impactaram no aprendizado da língua portuguesa, de modo que podemos defender, ainda que seja necessária uma análise mais próxima, que o processo de TLI teve atuação para a remarcação do parâmetro do sujeito nulo no português afro-brasileiro.

Longe de traçar definições ou tentar encerrar as intensas discussões em torno de um fenômeno tão estudado, foi possível perceber a força dos fatores estruturais para a atual configuração deste fenômeno variável, ao passo que as variáveis sociais não apresentaram uma relevância tão significativa (Duarte, 2019). A pesquisa linguística é, sobretudo, um trabalho descritivo. Logo, concluímos que ainda há muito o que se analisar em relação a um dos fenômenos “mais emblemáticos das contradições linguísticas” (Galves, 2019, p. 6).

REFERÊNCIAS

- ALMEIDA, Napoleão Mendes de. *Gramática Latina*. 26. ed. São Paulo: Saraiva, 1995.
- ALMEIDA, Napoleão Mendes de. *Gramática Metódica da Língua Portuguesa*. São Paulo: Saraiva, 2009.
- ALMEIDA, Norma Lucia Fernandes de. *Sujeito nulo e morfologia verbal no português falado por três comunidades do interior da Bahia*. 235 f. Tese (Instituto de Estudos da Linguagem) Universidade Estadual de Campinas, Campinas-SP, 2005.
- ALMEIDA, Norma Lucia Fernandes de.; CARNEIRO, Zenaide de Oliveira N. Sujeito. In: LOBO, T.; OLIVEIRA, K. (org.). *África à vista*. Salvador: EDUFBA, 2009. p. 50-69.
- ANDRADE, Manuel Correia de. *A terra e o homem no Nordeste*. São Paulo: Ciências Humanas, 1980.
- ANJOS, Rafael Sanzio Araújo dos. *Territórios das comunidades remanescentes de Antigos Quilombos no Brasil: primeira configuração espacial*. Brasília: Edição do autor, 1999.
- ARAÚJO, Maria do Socorro Targino de. *Os núcleos urbanos do recôncavo colonial*. 1970. Tese (Concurso de Assistente). Universidade Federal da Bahia, Salvador, 1970.
- ARAÚJO, Silvana Silva de Farias. *Nosso, da gente e de nós: um estudo sociolinguístico da expressão de posse no português rural afro-brasileiro*. 2005. (Dissertação). Universidade Federal da Bahia, Salvador, 2005.
- ARAÚJO, Silvana Silva de Farias. O português popular do semiárido baiano: fundamentos teóricos, sócio-históricos e empíricos. In: LOPES, Norma da Silva; OLIVEIRA, Josane Moreira; PARCERO, Lucia Maria de Jesus. *Estudos sobre o Português do Nordeste: língua, lugar e sociedade*. Editora Blucher, 2017, p. 45-72. Disponível em: <https://openaccess.blucher.com.br/article-details/o-portugues-popular-do-semiarido-baiano-20413>. Acesso em: 02.fev. 2022.
- ARAÚJO, Silvana Silva de Farias; ARAÚJO, Jean Marcel de Oliveira. *A formação sócio-histórica do português do Brasil: Contribuições do Recôncavo Baiano*. Cadernos de Letras da UFF – Dossiê: Difusão da língua portuguesa, n. 39, p. 95-116, 2009.
- ARAÚJO, Silvana Silva de Farias; ARAÚJO, Jean Marcel de Oliveira. Levantamento das comunidades rurais afro brasileiras do Recôncavo baiano: contribuições para futuros estudos sociolinguísticos em comunidades remanescentes de quilombos. In: FREITAG, Raquel Meister Ko.; SAVEDRA, Mônica Maria Guimarães. (Org.). *Mobilidades e Contatos Linguísticos no Brasil*. São Paulo: Editora Edgard Blücher Ltda., 2023, v. 1, p. 208-239.
- AZEVEDO, Aroldo de. *Vilas e cidades do Brasil colonial (ensaio de geografia urbana retrospectiva)*. Anais da Associação dos Geógrafos Brasileiros (1954-1955). São Paulo, v. 9, n. 1, p. 83-168, 1957.

BARICKMAN, Bert J. *Um contraponto baiano: açúcar, fumo, mandioca e escravidão no Recôncavo, 1780-1860*. Rio de Janeiro, Civilização Brasileira, 2003.

BAXTER, Alan.; LUCCHESI, Dante. Un paso más hacia la definición del pasado criollo del dialecto afro-brasileño de Helvécia (Bahia). In: ZIMMERMANN, K. (org.). *Lenguas criollas de base lexical española y portuguesa*. Frankfurt am Main: Vervuert: 1999. p. 119-141.

BENTLEY, W. Holman. *Dictionary and Grammar of the Kongo Language*. London, 1887.

BENVENISTE, Émile. A natureza dos pronomes. In: BENVENISTE, Émile. *Problemas de linguística geral* I. 4. ed. Campinas: Pontes, 1956.

BHAT, D. N. S. *Pronouns*. Oxford studies in typology and Linguistic Theory. New York: Oxford University Press, 2004.

BICKERTON, Derek. *Roots of Language*. Ann Arbor: Karoma Publishers, 1981.

BLOOMFIELD, Leonard. *Language*. New York, 1938, p.225-226.

BRASIL. Decreto nº 4.887, de 20 de novembro de 2003. Regulamenta o procedimento para identificação, reconhecimento, delimitação, demarcação e titulação das terras ocupadas por remanescentes das comunidades dos quilombos. *Diário Oficial da União*, Brasília, DF, 21 nov. 2003.

BRIGHT, William. (org.). (1966). *Sociolinguistics*. Haia: Mouton.

CASTILHO, Ataliba de. *Nova gramática do português brasileiro*. São Paulo: Contexto, 2010.

CHAMBERS, J. K. *Sociolinguistic theory*. Cambridge, Mass.: Blackwell Publishers, 1995.

CHATELAIN, Héli. *Grammatica Elementar do Kimbundu*. Genebra, 1888.

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

CHOMSKY, Noam; LASNICK, Howard. A Teoria de Princípios e Parâmetros. In: CHOMSKY, Noam. *O Programa Minimalista*. Tradução Eduardo Paiva Raposo. São Paulo: Editora Unesp, 2021[1993]. p. 57-214.

COELHO, Izete Lehmkuhl et al. *Para conhecer sociolinguística*. São Paulo: Contexto, 2015.

CUNHA, Ana Stela de Almeida. *A atuação do parâmetro do sujeito nulo na variedade popular rural do português do Brasil*. 2003. Tese (Doutorado em Linguística e Semiótica) – Universidade de São Paulo, São Paulo, 2003.

CUNHA, Ana Stela de Almeida. *Remanescentes de línguas crioulas em comunidades negras rurais isoladas: ausência de concordância de gênero na linguagem do Cafundó*. In: *ANAIIS DO GEL*. São Paulo, 1995.

CUNHA, Mário Pinto da. *História de Cruz das Almas*. Salvador: Imprensa Oficial da Bahia, 1959.

CYRINO, S. M. L. Observações sobre a mudança diacrônica no português do Brasil: objeto nulo e clíticos. In: ROBERTS, I.; KATO, M. A. (Org.). *Português brasileiro: uma viagem diacrônica: homenagem a Fernando Tarallo*. Campinas: Editora da Unicamp, 1993.

CYRINO, S. M. L.; DUARTE, M. E.; KATO, M. A. Visible subjects and invisible clitics in Brazilian Portuguese. In: KATO, M. A.; NEGRÃO, E. V. (eds.) *Brazilian Portuguese and the Null Subject Parameter*: 55-104. Frankfurt: Vervuert-Iberoamericana, 2000.

CYRINO, S.; MORAIS, M. A. T. (Coord.). *História do Português Brasileiro: mudança sintática do português brasileiro: perspectiva gerativista*. Coordenação geral: Ataliba T. de Castilho. São Paulo: Contexto, 2018.

DUARTE, M. E. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. In: ROBERTS, I.; KATO, M. A. (Org.). *Português brasileiro: uma viagem diacrônica. Homenagem a Fernando Tarallo*. Campinas: Editora da Unicamp, 1993. p. 107-128.

DUARTE, M; E. *A perda do princípio “evite pronome” no português brasileiro*. Orientadora: Profa. Dra. Mary A. Kato. Tese (Doutorado em Linguística). Universidade Estadual de Campinas, Campinas, 1995.

DUARTE, Maria Eugenia L. Termos da oração. In: VIEIRA, Silvia R. e Silvia F. BRANDÃO. 2007. *Ensino de Gramáticas: Descrição e Uso*. São Paulo: Editora Contexto, p. 262, 2007.

DUARTE, Maria Eugênia L.; MOURÃO, Gabriela C.; SANTOS, Heitor M. Os sujeitos de 3ª pessoa: revisitando Duarte. In: DUARTE, Maria Eugênia Lammoglia (Org.). *O sujeito em peças de teatro (1833-1992): estudos Diacrônicos*. São Paulo: Parábola, 2012, p. 21-44.

DUARTE, M. E. Do pronome nulo ao pronome pleno: a trajetória do sujeito no Português Brasileiro. In: ROBERTS, I.; KATO, M. A. (org.). *Português Brasileiro: uma viagem diacrônica. Homenagem a Fernando Tarallo*. 3ª ed. São Paulo: Contexto, 2018[1993]. p. 83-103.

DUARTE, M. E. O sujeito nulo referencial no Português Brasileiro e no Português Europeu. In: GALVES, C.; KATO, M. A.; ROBERTS, I. *Português brasileiro: uma segunda viagem diacrônica. Homenagem a Ilza Ribeiro*. Campinas: Editora da Unicamp, 2019a. p. 93-126.

DUARTE, M. E. *A remarcação em curso no valor do Parâmetro do Sujeito Nulo*. Cuadernos de la ALFAL, n. 12, v. 2, p. 71-99, noviembre 2020.

ECKERT, Penelope. Variation, convention and social meaning. Paper Presented at the *Annual Meeting of the Linguistic Society of America*. 2005. Oakland CA, Jan.7.

ECKERT, Penelope. Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation, 2012. *Annual Review of Anthropology*. Palo Alto.41:87-100.

FERNANDES, Henry Luydy Abraham. *As lâminas de machado lascadas Aratu de Piragiba-BA*. Salvador, 2011. 401f.

FERREIRA, Carlota. Remanescentes de um falar crioulo brasileiro. In: FERREIRA, Carlota et al. *Diversidade do português do Brasil*. Salvador: EDUFBA, 1984. p.21-32.

FIGUEIREDO SILVA, Maria Cristina. *A posição sujeito no português brasileiro: frases finitas e infinitivas*. Campinas: Ed. da UNICAMP, 1996.

FORCHHEIMER, Paul. *The category of person in language*. Berlin: Walter de Gruyter, 1953.

FREIRE, Gilberto. *Casa Grande e Senzala: formação da família brasileira sob o regime de economia patriarcal*. Rio de Janeiro: Record, 1997.

FREITAG, Raquel. *Problemas teórico-metodológicos para o estudo da variação linguística nos níveis gramaticais mais altos*. Matraca-Revista do Programa de Pós-Graduação em Letras da UERJ, v. 16, n. 24, 2009.

GALVES, Charlotte Marie Chambelland. *Ensaio sobre as gramáticas do português*. Campinas, SP: Editora da Unicamp, 2001.

GALVES, Charlotte Marie Chambelland. O que nos dizem as atas da Sociedade Protetora dos Desvalidos sobre a formação do português brasileiro (posfácio). In: GALVES, Charlotte; LOBO, Tânia (Orgs).: *O português escrito por afro-brasileiros no século XIX: as atas da Sociedade Protetora dos Desvalidos*. 1ª edição. Salvador: EDUFBA, 2019, p. 391-417).

GALVES, C.; KATO, M. A.; ROBERTS, I. *Português brasileiro: uma segunda viagem diacrônica*. Homenagem a Ilza Ribeiro. Campinas: Editora da Unicamp, 2019.

GUY, Gregory R. *Linguistic Variation in Brazilian Portuguese: Aspects of Phonology, Syntax and Language History*. University of Pennsylvania, PhD Dissertation, Ann Arbor, University Microfilms International, 1981.

GUY, Gregory R.; Ana Maria S. ZILLES, A. *Sociolinguística quantitativa – instrumental de análise*. São Paulo: Parábola Editorial, 2007.

HOLANDA, Sérgio Buarque de. *Raízes do Brasil*. Prefácio de Antônio Candido. 4. Ed. Brasília: Editora Universidade de Brasília, 1963.

HOLM, John. *Pidgins and Creoles. Volume 1: Theory and Structure*. Cambridge: Cambridge University Press, 1988.

HOLMBERG, Anders. *Is there a little pro? Evidence from Finnish*. *Linguistic Inquiry*, Cambridge, v. 36, n. 4, p. 533–564, 2005.

HOUAISS, A. (1985). *O português do Brasil*. Rio de Janeiro: UNIBRADE/UNESCO.

HUANG, J. *On the distribution and reference of empty pronoms*. *Linguistic Inquiry*, 531p.1984.

JAEGGLI, O. & SAFIR, K. The null subject parameter and parametric theory. In: O. Jaeggli & K. Safir (eds). *The Null Subject Parameter*. Dordrecht, Kluwer, 1987.

KATO, Mary Aizawa. The partial pro-drop nature and the restricted VS order in Brazilian Portuguese. In: KATO, M. A.; NEGRÃO, E. V. (ed.). *Brazilian Portuguese and the null subject parameter*. Madrid: Vervuert, Iberoamericana., 2000. p. 223-258.

KATO, Mary. *A evolução da noção de parâmetros*. DELTA: Documentação de Estudos em Linguística Teórica e Aplicada, v. 18, p. 309-337, 2002.

KOWARICK, Lúcio. *Trabalho e vadiagem: a origem do trabalho livre no Brasil*. São Paulo: Brasiliense, 1987.

KURYŁOWICZ, Jerzy. *Esquisses linguistiques*. Wrocław: Zakład Narodowy im. Ossolińskich, 1960.

LABOV, William (1972) *Sociolinguistics patterns*. Philadelphia: University of Pennsylvania Press.

LABOV, William. *Padrões sociolinguísticos*. Tradução de Marcos Bagno, Maria Marta Pereira Scherre e Caroline Rodrigues Cardoso. São Paulo: Parábola Editorial, 1994 [2008].

LABOV, William. *Quantitative reasoning in linguistics*. Disponível em: <https://www.ling.upenn.edu/~wlabov/Papers/QRL.pdf> Acesso em: 25 jan. 2024.

LAVANDERA, Beatriz. *Where does the sociolinguistic variable stop?* *Language in society*, v. 7, n. 2, p. 171-182, 1978.

LEÃO, Sônia de Oliveira. *Padrões de desenvolvimento urbano (1500-1930)*. In: SILVA, S. C. B. de M.; LEÃO, S. de O.; SILVA, B-C. N. (Org.). *Urbanização e metropolização no Estado da Bahia: evolução e dinâmica*. Salvador: Centro Editorial e Didático da UFBA, 1989.

LIGHTFOOT, David. *How to set parameter: Arguments from language change*.

Cambridge, MA: MIT Press. 1991.

LINHARES, Maria Yeda (Org.). *Brasil: História Econômica*. Rio de Janeiro: Editora Elsevier, 2008.

LOBO, Tania Conceição Freire. A linguística histórico-diacrônica no Brasil pós-1980 e a questão do contato linguístico. *Diadorim*, Rio de Janeiro, vol. 20 – Especial, p. 64–80, 2018.

LOPES, Célia R. dos Santos. *A inserção de 'a gente' no quadro pronominal do português*. Frankfurt/Madri: VERVUERT/Iberoamericana, 2003.

LOPES, Célia; BROCARD, Maria Teresa. Main Morphosyntactic Changes and Grammaticalization Processes. In: WETZELS, W. Leo.; MENUZZI, Sergio; COSTA, João (Ed.). *The Handbook of Portuguese Linguistics*. Malden/Oxford: Wiley Blackwell, 2016, p. 471-486.

LOREGIAN-PENKAL, Loremi. *(Re)Análise da Referência de Segunda Pessoa na Fala da Região Sul*. 2004. 260f. Tese (Doutorado em Letras) – Universidade Federal do Paraná, Curitiba, 2004.

LUCCHESI, Dante. The article systems of Cape Verde and São Tomé Creole Portuguese: general principals and specific factors. *Journal of Pidgin and Creole Languages*, Filadélfia; Amsterdam, v.8, n.1, p.81-108, 1993.

LUCCHESI, D. *Sistema, Mudança e Linguagem: um percurso na história da língua*. São Paulo: Parábola, 2004.

LUCCHESI, D.; BAXTER, A.; RIBEIRO, I. (org.). *O português afro-brasileiro*. Salvador: EDUFBA, 2009.

LUCCHESI, Dante. *Língua e sociedade partidas: a polarização sociolinguística do Brasil*. São Paulo: Contexto, 2015.

LUCCHESI, Dante. A Teoria da Variação Linguística: um balanço crítico. *Estudos Linguísticos*, São Paulo, v. 41, n. 2, p. 793-805, 2012.

LUCCHESI, Dante Eustachio; SOUZA, Jurgen Alves de. A variação no uso do pronome reflexivo no português afro-brasileiro como resultado de mudanças induzidas pelo contato entre línguas no passado. *Diadorim: Revista de Estudos Linguísticos e Literários*, Rio de Janeiro, v. 20, número especial, p. 458–488, 2018. DOI: <https://doi.org/10.35520/diadorim.2018.v20n0a23286>.

LUCCHESI, Dante. *Por que a crioulação aconteceu no Caribe e não no Brasil: condicionamentos sócio-históricos*. Gragoatá, Niterói, v.24, n. 48, p. 227-255, jan.-abr. 2019.

MALTA, Daniela Paula de Lima Nunes. *A natureza da morfologia de flexão verbal e o parâmetro do sujeito nulo em dados escritos de variedades africanas do português: uma análise contrastiva*. 2019. 237 f. Dissertação (Mestrado em Letras) – Universidade Federal de Pernambuco, Recife, 2019.

MALTA, Daniela Paula de Lima Nunes; SILVA, Cláudia Roberta Tavares. *O sujeito nulo na variedade africana do português guineense: expressão e morfologia de flexão verbal em dados escritos*. Anais do CONEIL, 2020.

MAGALHÃES, Telma Moreira Vianna. *O sistema pronominal sujeito e objeto na aquisição do português europeu e do português brasileiro*. 2006. 175 f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2006.

MARTIN, Gabriela. *Pré-História do Nordeste do Brasil*. 3ª Ed. Recife: Edit. Universitária UFPE, 1999.

MARTINS, Marco Antonio; COELHO, Izete Lehmkuhl; CAVALCANTE, Sílvia Regina de Oliveira. Variação sintática e gerativismo. In: Marco Antonio Martins; Jussara Abraçado. (Org.). *Mapeamento Sociolinguístico do Português Brasileiro*. 1ed. São Paulo: Contexto, 2015, v, p. 221-248.

MATEUS, Maria Helena M. et al. *Gramática da Língua Portuguesa*. Caminho Lisboa, 2003.

MATTOS E SILVA, Rosa Virgínia. *Ensaio para uma sócia-história do português brasileiro*. São Paulo: Parábola Editorial, 2004.

MATTOSO, Kátia Maria de Queirós. *Bahia: a cidade de Salvador e seu mercado no século XIX*. São Paulo: HUCITEC; Salvador: Secretaria Municipal de Educação e Cultura, 1978. p. 155-169.

MEILLET, Antoine. (1906). Comment les mots changent de sens. *L'Année sociologique* (1905-1906); reimpresso in *Linguistique historique et linguistique générale*. I, 230-271.

MONTEIRO, José Lemos. *Pronomes pessoais: subsídios para uma gramática do português do Brasil*. Fortaleza: Edições UFC, 1994.

NARO, Anthony J. “Modelos Quantitativos e Tratamentos Estatísticos” in MOLLICA, M.C. (org.). *Introdução à Sociolinguística*. Rio de Janeiro: Cadernos Didáticos da UFRJ, 1994 (p 17/25).

MOURA, Glória. A força dos tambores: a festa nos quilombos contemporâneos. In: SHWARCZ, Lília Moritz; REIS, Letícia Vidor de Souza (Org.). *Negras Imagens: ensaios*

sobre cultura e escravidão no Brasil. São Paulo: Ed. Universidade de São Paulo: Estação Ciência, 1996.

MUNANGA, Kabengele. Identidade, cidadania: algumas reflexões sobre os discursos anti-racistas no Brasil. In: QUINTAS, Fatima (Org.) *O negro: Identidade e cidadania*. Anais IV Congresso Afro-Brasileiro, Fundação Joaquim Nabuco, Recife, Editora Massangana, 1995, p. 66.

MUSSA, Alberto. (1991). *O papel das línguas africanas na história do português do Brasil*. Dissertação de mestrado, UFRJ, mimeo.

NARO, A. J.; LEMLE, M. Syntactic diffusion. In: STEEVER, S. B. et al (eds.). *Papers from the parasession on Diachronic Syntax*. Chicago: Chicago Linguistic Society, 1976. 221-241.

NARO, Anthony Julius; SCHERRE, Maria Marta Pereira (1993): *Sobre as origens do português popular do Brasil*. DELTA. São Paulo, Educ., 9. p. 437-454.

NICOLAU, E. (1995). *As propriedades de sujeito nulo e ordem V-S no português brasileiro*. Tese de doutorado. Campinas: UNICAMP.

NURSE, Derek; PHILIPPSON, Gérard (org.). *The Bantu Languages*. London: Routledge, 2003.

OMENA, N. P. *Pronome pessoal de terceira pessoa: suas formas variantes em função acusativa*. Rio de Janeiro: PUC-RJ, 1978. Dissertação (Mestrado em Língua Portuguesa) do Departamento de Letras, Pontifícia Universidade Católica, PUC, Rio de Janeiro, 1978.

PAGOTTO, Emilio Gozze. *Sociolinguística*. In: Introdução às Ciências da Linguagem – Linguagem, História e Conhecimento. J. H. Nunes e C. C. Pfeiffer (Orgs.). Campinas: Fontes, 2006, p. 49-72.

PARKVALL, Mikael.; LÓPEZ, Laura A. *Português vernáculo brasileiro e a hipótese da semi-crioulização*. Revista da ABRALIN, n. 2, v. 1, p. 111-152, 2003.

PEROTA, Celso. *Considerações sobre a cerâmica Aratu nos Estados da Bahia e Espírito Santo*. Boletim do Museu de História e arte - Arqueologia nº 1. Vitória: UFES, 1971.

PETTER, Margarida. O português do Brasil: contato linguístico, história e diversidade. In: LUCCHESI, Dante; BAXTER, Alan; RIBEIRO, Ilza (orgs.). *O português afro-brasileiro*. Salvador: EDUFBA, 2008. p. 35–60.

PITOMBO, Eliana; ALMEIDA, Norma. O sujeito nulo. In. GALVES, Charlotte; LOBO, Tânia (Orgs.): *O português escrito por afro-brasileiros no século XIX: as atas da Sociedade Protetora dos Desvalidos*. 1ª edição. Salvador: EDUFBA, 2019, p. 259-289.

PRADO JR., Caio. *Formação do Brasil contemporâneo*. São Paulo: Brasiliense, 1999. p. 306.

PRATAS, F. Expletive *pro* and overt verbal agrément morphology: (Conter-) evidence from Capeverdean. In: *14th Colloquium on Generative Grammar*, 2004, Porto. Handout, Porto, 2004, p. 1-14.

PTDRSS. *Plano de Desenvolvimento Rural, Sustentável e Solidário do Recôncavo*. SABLAYROLLES, P. J. L.; VELLOSO, T. R.; JESUS, C. S. de. (orgs). Cruz das Almas-BA: UFRB/NEDET, 2017.

PULGRAM, Ernst Statics and Dynamics of Linguistic Subcodes. *Lingua*, 1961.

REIS, João José. *Liberdade por um fio: história dos quilombos no Brasil*- São Paulo: companhia das letras, 1996.

RIBEIRO, Ilza. *A mudança sintática do português brasileiro é mudança em relação a que gramática?* In: *Para a história do português brasileiro*. Vol. I – Primeiras ideias, 1998, p. 101-119.

RIZZI, Luigi. *Issues in Italian syntax*. Dordrecht: Foris, 1982.

ROBERTS, I. Gramáticas “marginais” e mudanças sintáticas “extremas”: o inglês e o português brasileiro. In: GALVES, C.; KATO, M.; ROBERTS, I. *Português brasileiro: uma segunda viagem diacrônica*. Homenagem a Ilza Ribeiro. Campinas: Editora da Unicamp, 2019. p. 23-56.

ROBERTS, I.; KATO, M. (Org.). *Português brasileiro: uma viagem diacrônica*. São Paulo: Contexto, 2018

RODRIGUES, A. (1987). *Concordância verbal no dialeto popular da cidade de São Paulo*. Tese inédita, São Paulo: USP.

SANKOFF, David. *Sociolinguistics and syntactic variation*. In: *Linguistics: the Cambridge survey*. Frederick J. Newmeyer (ed.). Cambridge: Cambridge University Press, 1988.

SCHERRE, M. M. P. 2002. A norma do imperativo e o imperativo da norma– Uma reflexão sociolinguística sobre o conceito de erro. In: M. BAGNO (org.) *Linguística da norma*. São Paulo: Loyola. 217-251.

SCHWARTZ, Stuart. *Escravos, roceiros e rebeldes*. Bauru/SP: EDUSC, 2001. p. 123-170.

SILVA, Claudia Roberta Tavares. *A natureza de AGR e suas implicações na ordem VS: um estudo comparativo entre o português brasileiro e o português europeu*. 2004. Tese (Doutorado em Linguística). Universidade Federal de Alagoas, Maceió.

SILVA, Cláudia Roberta Tavares; ZIOBER, Fernanda Maciel. *Sobre os sujeitos pré-verbais duplicados: uma análise contrastiva entre o português vernacular brasileiro, o caboverdiano e o santome*. In: *Estudos Linguísticos e literários*, Nº 57, jul-dez|2017, Salvador: p. 164-185.

Disponível em:

<<https://periodicos.ufba.br/index.php/estudos/article/download/24295/15722/90000>>

Acesso em: 07 mai 2024.

SILVA, Jacson Balduino Silva; ARAUJO, Silvana Silva de Farias; SANTIAGO, Huda da Silva. *História social e mudança linguística: sujeito pronominal, concordância verbal e contatos linguísticos. No prelo.*

SILVA, Jacson Balduino. *Efeitos da mudança sintática na realização do sujeito pronominal no português da comunidade quilombola Mussuca*. 131f. Dissertação (Programa de Pós-graduação em Estudos Linguísticos) – Universidade Estadual de Feira de Santana, Feira de Santana-BA, 2023.

SILVA, Jacson Balduino Silva; ARAUJO, Silvana Silva de Farias. *Aquisição de linguagem, contatos linguísticos e remarcação paramétrica no português do Brasil*. Organon. Porto Alegre, v. 38, n. 76, jul/dez. 2023.

SILVA, Pablo Rezende da. *Entre quilombos: uma história de luta, resistência e de um modo de vida comunitária*. 104f. Dissertação (Programa de Pós-graduação Profissional em Educação do Campo) – Universidade Federal do Recôncavo da Bahia, Amargosa-BA, 2020.

TAGLIAMONTE, Sali A. 2006. *Analysing sociolinguistic variation*. New York: Cambridge University Press. 284p.

TARALLO, Fernando. Diagnosticando uma gramática brasileira: o português d'aquém e d'além-mar ao final do século XIX. In: ROBERTS, Ian; KATO, Mary Aizawa (orgs.). *Português brasileiro: uma viagem diacrônica*. Campinas: Editora da UNICAMP, 1993. p. 69-105.

TARALLO, Fernando. Sobre a alegada origem crioula do português brasileiro: mudanças sintáticas aleatórias & Diagnosticando uma gramática brasileira: o português d'aquém e d'além-mar no final do século XIX, In: KATO, Mary A. & ROBERTS Ian. *Português brasileiro: uma viagem diacrônica*. Campinas: Editora da UNICAMP, 1996.

TARALLO, F. Sobre a alegada origem crioula do Português Brasileiro: mudanças sintáticas aleatórias. In: ROBERTS, I.; KATO, M. A. *Português brasileiro: uma viagem diacrônica*. São Paulo: Contexto, 2018[1993]. p. 29-53.

TEYSSIER, Paul. *História da língua portuguesa*. Tradução de Celso Cunha. São Paulo: Martins Fontes, 2007.

VEIGA, Manuel. *O crioulo de Cabo Verde: introdução à gramática*. Praia: Instituto Cabo-Verdiano do Livro, 1995.

WEINREICH, Uriel.; HERZOG, Marvin; LABOV, William. *Fundamentos empíricos para uma teoria da mudança linguística*. Trad.: M. Bagno. São Paulo: Parábola Editorial, 2006.

APÊNDICE - ROTEIRO DA ENTREVISTA

ROTEIRO PARA O PROJETO DE PESQUISA “O USO DO PRONOME NULO E DO PRONOME PLENO NA POSIÇÃO SUJEITO NA LÍNGUA FALADA DA COMUNIDADE BAIXA DA LINHA NO RECÔNCAVO BAIANO”

UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

A entrevista sociolinguística não é uma sequência de perguntas em um questionário, pois quem realiza a entrevista deve ter a liberdade para mudar o encaminhamento das perguntas de modo a permitir a fala sobre diferentes temas.

Com a exceção da checagem com base na ficha social (nome, idade, onde mora etc.) e do protocolo de atitudes e percepções linguísticas, o roteiro da entrevista abarcou narrativas referentes a temas relacionados à infância, temas relacionados a risco de vida, mas também argumentações e explanações sobre temas relacionados a estudo, trabalho, política, cidade, meio ambiente, culinária, entre outros, além de perguntas sobre a comunidade (gosta/não gosta; quer ficar; o que pensa) e perguntas de avaliação da fala. Os cinco minutos finais da entrevista serão dedicados ao protocolo de percepções e atitudes linguísticas, ou seja, os julgamentos do participante ao correlacionar fatores sociais a traços linguísticos e seus pensamentos sobre os fatos da língua: uso variável da concordância nominal e verbal de número, de preposições, de expressão de sujeitos pronominais de segunda pessoa do singular (*tu ~ você*) e de 1ª pessoa do plural (*nós ~ a gente*), o uso variável da flexão de caso nos pronomes, objeto direto anafórico, entre outros.

ANEXOS

PROJETO
VERTENTES DO PORTUGUÊS POPULAR DO ESTADO DA BAHIA

CHAVE DE TRANSCRIÇÃO

DANTE LUCCHESI

1. Apresentação e objetivos

Esta proposta de chave de transcrição tem orientado os trabalhos de transcrição ortográfica dos inquéritos realizados no âmbito do Projeto Vertentes do Português Popular do Estado da Bahia. Optou-se pela transcrição ortográfica em detrimento da fonética, porque aquela possibilita uma melhor visualização do texto, atendendo igualmente às necessidades da análise. Contudo, diferentemente do procedimento adotado em outros projetos (NURC, por exemplo), a transcrição ortográfica é aqui adotada como um instrumento efetivo para a análise, ou seja, ela é feita no sentido de que possa fundamentar a coleta dos dados a serem analisados, diminuindo o recurso à gravação. Desse modo, a transcrição deve mobilizar os recursos disponíveis para uma reprodução mais fiel das características linguísticas do texto oral.

2. Concepção e considerações prévias

O critério básico é o de se registrarem os fatos linguísticos que se observam na fala do informante³⁰ e que constituem marcas específicas do seu dialeto, tanto no nível fônico, quanto no morfo sintático (ausência de concordância ou de nexos gramaticais - preposições, complementizadores, etc. -, objetos nulos, quebras no encadeamento sintático dos enunciados, etc.); serão excluídos apenas alguns fatos que, mesmo em desacordo com a ortografia oficial,

³⁰ bem como do documentador, para a possível consideração do "efeito de gatilho" na análise. O documentador por vezes aproxima a sua fala da do informante, como procedimento técnico de recolha do material, e não deve ser "corrigido" na transcrição.

podem ser considerados gerais no português do Brasil, com o intuito de não sobrecarregar o texto transcrito.

Esta chave de transcrição está organizada da seguinte maneira. Na seção 3, são feitas algumas indicações preliminares sobre a formatação do texto transcrito. Na seção 4, são definidos alguns critérios gerais para orientar todo o trabalho de transcrição. Na medida em que o princípio geral é o de registrar todos os fatos da fala do informante, será feito primeiramente um destaque para os fatos que não devem ser registrados. Esse procedimento visa a facilitar o trabalho de transcrição, pois ao transcrever o pesquisador só terá em mente um número reduzido de fatos que não devem ser registrados. Os demais fatos devem, em princípio, ser registrados. Para esses fatos, como subsídio, serão apresentadas, na seção 5, algumas indicações mais específicas de como eles devem figurar na transcrição dos inquéritos. Obviamente que esta última listagem não contempla toda a gama de fatos que ocorrerão nos inquéritos. Assim, os fatos novos deverão ser tratados segundo o procedimento geral desta chave, devendo ser comunicados ao seu autor para que a solução adotada seja ratificada e incorporada.

3. Indicações preliminares

1. Cabeçalho: modelo:

PROJETO VERTENTES DO PORTUGUÊS RURAL DO ESTADO DA BAHIA
LOCALIDADE DE HELVÉCIA-BA

INFORMANTE 01

NOME: - SEXO: F IDADE:
NASCIMENTO: - PAIS: Helvécia
NIVEL DE ESCOLARIDADE: ANALFABETA
VIAGENS PARA FORA: NÃO

2. Os intervenientes no inquérito devem ser indicados da seguinte maneira:

DOC: (documentador)

INF: (informante)

CIRC: (interveniente circunstancial)

3. Indicação de trecho ininteligível: ININT

4. Indicação de interrupção do inquérito: INTERRUP

5. O texto deve ser pontuado com moderação.

6. Usar colchetes para indicar trecho sobre qual não há certeza na audição.

Ex.: INF: Ah, pra passeá, [eu] saio.

7. Indicação de realização inusitada de uma palavra ou expressão em itálico.

Ex.: Então eu *abandonei* a escola.

OBS.: sem o itálico o leitor tende a pensar que o digitador esqueceu o “n”, ao invés de aceitar a realização inusitada.

8. Palavra desconhecida, ouvida nitidamente na transcrição/revisão, também deve ser grifada em itálico.

Ex.: ... levava água no *carote*.

9. Se o documentador fizer algum comentário durante a fala do informante, sem interromper o seu fluxo sintático-discursivo, deve-se indicar com reticências a continuidade desse fluxo do informante, intercalando a fala do documentador.

Ex.: INF: Naquela época, eu saía muito...

DOC: Ah, legal!

INF: ... com os camarada pras festa.

10. Nos casos de discurso direto, deve-se utilizar aspas e pontuar antes de fechar as aspas.

Ex.: Ele disse: “Eu vô embora. Já tá tarde.” Aí eu falei: “Vá não. Durma aqui em casa.”

11. recursos não verbais empregados pelo informante, entre chaves:

Ex.: Quando era, assim, perto de mei’ dia, assim, pai começava a ‘subiá: *{o informante assovia}*.

4. Critérios Gerais

1. **Deverão ser registrados, com o máximo de rigor atenção, TODOS os fatos no nível da morfossintaxe, tais como:**

a) concordância nominal variável:

Ex.: *as coisa tá caro.*

estava no roça.

b) concordância verbal variável.

Ex.: *os menino foi* ou *os menino foro*; *nós vai*; ou mesmo *eu conta por eu conto.*

c) Omissão de preposições, artigos, complementizadores, etc.

Ex.: Deu Júlia isso aí.

(Port. Padrão: Deu à Júlia isso aí.)

d) repetições, com utilização das reticências.

Ex.: As pessoas assim é que... que conhece mais as coisa.

e) correções e quebras no encadeamento do enunciado, também com a utilização de reticências.

Ex.: E as... sempre assim ou era diferente?

f) hesitações, idem.

Ex.: Esses daqui mesmo, ói, é... é... é... é d'uma... de... de Odete aqui, num sabe?

g) separar o tópico por vírgula

Ex.: Aqueles ossozim pequeno, era os bezerro.

Aqueles de mais de longe, vamo soltá.

h) a dupla negação deve vir sem vírgulas.

Ex.: Não vi ele não.

i) deve-se distinguir o *né* de valor estritamente fático, que deve vir entre vírgulas, do *né* com entonação de pergunta, o qual deve vir acompanhado por ponto de interrogação.

2. Fatos fônicos que não devem ser registrados

a) Elevação das vogais médias em distribuição pré-acentuada, quer no interior de um mesmo vocábulo, quer no interior de um grupo de força (como no caso das seqüências com a preposição **em**), e pós-acentuada não-final;

Ex: *estava*, e não *istava*; *domingo*, e não *dumingo*; *em casa*, e não *im casa*; *rédea*, e não *rédia*.

Exceção: transcrever *ni* e não *ne*: *fui ni Feira de Santana*.

b) Elevação das vogais médias em posição átona final.

Ex.: [nomi] > *nome*; [matu] > *mato*

c) Ditongação antes de consoante constrictiva implosiva.

Ex.: [meys] > *mês*

d) Ditongação da terminação nasal -em

Ex.: *tem*, e não *teim*; *ninguém*, e não *ninguéim*.

e) Palatalização de **t** e **d**, antes de vogal palatal.

Ex.: *ponte*, e não *pontche*; *noite*, e não *noitche*.

Obs.: palatalizações mais desviantes da norma urbana devem ser registradas: *mntcho*, *dódio*, *tchem*, *tcherra*.

f) epêntese da vogal alta que desfaz o travamento de grupo consonântico.

Ex.: *advogado*, e não *adivogado*; *psicologia*, e não *pissicologia*.

Obs.: A inserção de uma vogal média deve ser registrada: *adevogado*.

g) Realização velar ou faríngea de certas consoantes constrictivas:

Ex.: *tava* por [ˈtaha] ou [ˈtaxa]; *mais* por [mayh]

h) Supressão da consoante nasal

Ex.: *tenho* por [teⁿyu]; *uma* por [uⁿa]

i) Vocalização da consoante lateral pós-vocálica

Ex.: *mel*, e não *méu*; *Brasil*, e não *Brasiu*; *animal*, e não *animau*.

5. Detalhamento

1. Consoantes implosivas – (C)VC

a) /S/

- a variante zero deve ser registrada: *os menino*.
- as variantes velar/glotal não devem ser registradas: *mesmo* para [‘mehmo].

b) /l/

- a variante zero deve ser registrada: *papé, animá, horríve* etc.
- as variantes constrictivas, [x] e [h], devem ser registradas com o grafema <r>:

vorta, arcançou, etc.

c) /x/

- a variante zero deve ser registrada: *brincá, corrê, dormí, partí*.

Exceção: A forma do verbo *ir* deve ser sempre transcrita como *ir*.

- a variante [w] deve ser registrada: *felvendo, nelvoso*.

d) /N/

- deve-se registrar a redução dos ditongos nasais em final de palavra: *viage,*

bença, correro.

e) A inserção de uma vogal, criando uma nova sílaba (CVC > CVCV):

dificulidade; tale, capinho.

f) semivogais

- Redução de ditongos, tanto em posição final, quanto em posição medial, utilizando o acento circunflexo na sílaba tônica.

Ex.: *matô; côro; poquinho; cantê; dêxa.*

Exceção: a conjunção *ou* deve ser sempre transcrita por *ou* e não *ô*.

2. Padrão silábico CCV

- redução do grupo (CCV > CV): *cumpade, ôto, péda.*
- CCV > CVC: *percurar.*
- CVC > CCV: *preguntô; drumença.*
- CCV > CVCV: *fulô.*

3. Vogais

- Abaixamento das vogais altas em distribuição não acentuada.

Ex.: *deferente* por *diferente*.

b) Redução específica de certos ditongos como em *truxe* [trusi] por trouxe, em que ocorre também a elevação da vogal.

c) Nasalização.

Ex.: *ingual* por igual; *ingnorância* por ignorância, *inzame* por exame (neste caso, colocar em itálico).

d) Usa-se o acento de timbre para indicar uma realização distinta da normal.

Ex.: *interésse* por interesse.

e) Troca de *e* ou *en* por *a* ou *an*, e vice-versa.

Ex.: *antão* por então.

f) Troca de vogal posterior pela anterior, em distribuição átona final.

Ex.: *quande* por quando; *devide* por devido; *tem* por tenho.

g) Ditongação.

Ex.: *saudia* por sadia.

4. Consoantes

a) o "ieísmo".

Ex.: *muié* por mulher; *véizim* por velhinho; *trabaio* por trabalho.

b) o rotacismo.

Ex.: *pranta* por planta; *prantação* por plantação.

5. Inserção e supressão de segmentos fônicos

a) inserção de segmento no início do vocábulo.

Ex.: *ieu* por eu; *amontá* por montar; *evém* por vem.

b) supressão de segmento no início do vocábulo, com apóstrofo

Ex.: *'bservano* por observando; *'djutóro* por adjutório; *'cabô* por acabou; *'fraquiceno* por enfraquecendo; *'ocê* por você.

Exceção: *cê, tá, tão, tamos*.

c) supressão de segmento no final do vocábulo, com a utilização do apóstrofo quando ocorrer junção com vocábulo posterior.

Ex.: *quan' nada*; *den' de*; *des' tamanho*; *pó' dêxá*.

d) a elisão entre palavras diferentes.

Ex.: *dêx'eu*; *pr'aculá*; *s'istrompô*; *mand'ele*.

Exceção: a elisão com o *que* não deve ser registrada; *que eu* e não *qu'eu*; *que é* e não *qu'é*.

e) redução de terminações verbais.

Ex.: *cantaro* ou *cantarum* por cantaram; *brincano* por brincando.

f) redução do morfema de diminutivo.

Ex.: *véim* por velhinho; *urim* por ourinho.

g) redução do ditongo /ãw/.

Ex.: *Som Paulo* por São Paulo.

h) redução de proparoxítonas, bem como das falsas proparoxítonas, com a manutenção do acento para facilitar o entendimento da forma, quando necessário.

Ex.: *épa* por época; *lâmpa* por lâmpada; *abróba* por abóbora; *estâumbo* por estômago; *remédo* por remédio; *criatóro* por criatório; *muléstra* por moléstia.

i) A combinação de fatos acima mencionados, com a utilização do acento de timbre para evitar ambigüidade, quando for o caso.

Ex.: *quês* por aqueles; *véi* por velhos (para distinguir *véi* de *vei* – forma reduzida de veio) ; *fî* por filho; *ói* por olhe.

j) aglutinação.

Ex.: *nestante*; *vumbora*, *vou m'embora*.

l) as interjeições são sempre grifadas com h:

Ex.: *eh...* e não *é...*; *ah...*; *ham...*; *ham-ham*; *hum-hum*; *hum...*; *oh*; *ôh*; *hei*.