



Universidade Federal da Bahia
Instituto de Letras
Programa de Pós-Graduação em Língua e Cultura
Rua Barão de Jeremoabo, nº147 - CEP: 40170-290 - Campus Universitário Ondina Salvador-BA
Tel.: (71)3283 - 6256 – Site: <http://www.ppglinc.letas.ufba.br/> - E-mail: pgletba@ufba.br

GRACIELLE DE BARROS JESUS

**A INDETERMINAÇÃO DO SUJEITO NO PORTUGUÊS POPULAR DE
SALVADOR**

SALVADOR

2017

GRACIELLE DE BARROS JESUS

**A INDETERMINAÇÃO DO SUJEITO NO PORTUGUÊS POPULAR DE
SALVADOR**

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

Orientador: Prof. Dr. Alan Norman Baxter

SALVADOR

2017

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

GRACIELLE DE BARROS JESUS

**A INDETERMINAÇÃO DO SUJEITO NO PORTUGUÊS POPULAR DE
SALVADOR**

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

Aprovada em 28 de março de 2017.

Banca Examinadora:

Maria Cristina Vieira de F. Silva
Doutora em Letras, PPGLinC
Universidade Federal da Bahia (UFBA-Presidente da Banca Examinadora)

Marcela Paim
Doutora em Letras, PPGLinC
Universidade Federal da Bahia (UFBA)

Vivian Antonino da Silva
Doutora em Letras
Universidade Estadual do Sudoeste Baiano (UESB)

Sistema de Bibliotecas da UFBA

de Barros Jesus, Gracielle
A indeterminação do sujeito no português popular de Salvador
/ Gracielle de Barros Jesus. -- Salvador, 2017.
120 f.

Orientador: Alan Baxter.
Dissertação (Mestrado - Letras) -- Universidade Federal da
Bahia, UFBA, 2017.

1. Indeterminação do sujeito. 2. Português popular . 3.
Projeto Vertentes. 4. Transmissão Linguística Irregular . I.
Baxter, Alan. II. Título.

AGRADECIMENTOS

Todos nós sabemos que, por mais que a jornada acadêmica pareça, às vezes, solitária e nos afaste de algumas pessoas, nós não somos uma ilha, sendo que o apoio daqueles que nos cercam é fundamental para que as coisas se tornem menos difíceis. Assim, me proponho, através dessas singelas, mas muito sinceras palavras, demonstrar minha eterna gratidão àqueles que estiveram do meu lado durante todo esse tempo.

Ao professor Alan Baxter, pela oportunidade e pela confiança, além da paciência e apoio nos momentos difíceis dessa jornada;

Ao professor Dante Lucchesi, por me apresentar ao Projeto Vertentes e à Sociolinguística, e por confiar no meu potencial desde a Iniciação Científica;

Aos meus pais Antonio e Kadja e à minha irmã Adrielle, por não me negarem apoio e amor em nenhum momento da carreira acadêmica e da vida;

A Danilo, cujo amor e carinho me mantêm firme diante de todas as intempéries, e por sempre confiar mais em mim do que eu mesma;

A minha Elis, que, apesar de ser uma realidade recente, me fez esquecer de tudo o que veio antes. Todo meu esforço tem você como maior motivação, minha filha!;

Aos meus compadres e amigos que tanto amo Michele e Vinícius, por terem sido tão pacientes durante as minhas crises;

A minha querida Carla, que me fez rir com suas “tretas” mesmo quando eu queria chorar de desespero;

Aos colegas do Projeto Vertentes, pelo acolhimento e companheirismo, com um “obrigada” especial a Lanuza, pelo apoio e pelas revisões tão valorosas;

A todos os professores, com quem aprendi muito durante as etapas de cumprimento de créditos;

Aos queridíssimos amigos que o mestrado trouxe para a minha vida: Jane, Angelo, Aline, Lorena, Danilo, Sebastian, Ingrid, e aos que já estavam aqui, Elias e Elaine;

A Cezar, que merece um agradecimento individual (rsrs) por ter sido tão importante pra mim desde o primeiro momento que nos conhecemos. Amizade à primeira vista também existe e nós somos a prova disso!;

Aos meus alunos queridos de LETE43, por terem me provado que a docência foi a melhor escolha que eu poderia ter feito;

Aos informantes, sem os quais essa pesquisa não teria sido possível;

À UFBA, pela excelência;

À FAPESB, por ter financiado essa pesquisa.

*“Dá-me um beijo”, ela me disse,
E eu nunca mais voltei lá.
Quem fala ‘dá-me’ não ama,
Quem ama fala ‘me dá’
‘Dá-me um beijo’ é que é correto,
É linguagem de doutor,
Mas ‘me dá’ tem mais afeto,
Beijo me-dado é melhor.”*

(Mário Lago)

RESUMO

Nessa dissertação, estudaremos as estratégias pronominais para expressar a indeterminação do sujeito no português popular da Bahia, estudando amostras de fala de quatro bairros da cidade de Salvador: Liberdade, Cajazeiras, Itapuã e Subúrbio. Traremos, aqui, os resultados obtidos a partir da análise das variáveis sociais nas quais os informantes estão estratificados, e seguiremos a hipótese de que o quadro atual de variação das estratégias pronominais de indeterminação é resultado de mudanças desencadeadas pelo contato do português com as línguas indígenas e africanas, nos primeiros séculos da formação da sociedade brasileira. Observaremos, também, se existem diferenças significativas entre os resultados encontrados nas localidades. As análises aqui realizadas seguem os pressupostos teóricos e metodológicos da Sociolinguística Variacionista (LABOV, 1972) e nossas hipóteses se relacionam diretamente com o processo de Transmissão Linguística Irregular (BAXTER, 1995; LUCCHESI & BAXTER, 2009) e da realidade bipolarizada do português brasileiro (LUCCHESI, 2015). Na nossa análise, foram identificadas cinco estratégias pronominais de indeterminação do sujeito: a gente, você, eles, nós e Ø + verbo na terceira pessoa do singular (Ø + V3PS). Essas variantes foram analisadas de acordo com as variáveis linguísticas Realização fonética do sujeito indeterminado, Desinência verbal, Paralelismo formal, Referência ao falante, Tipo de verbo, Tipo de frase, Nível de referencialidade do agente e Modo, e com as variáveis sociais Sexo, Faixa etária, Estado fora da comunidade, Escolaridade e Localidade.

Palavras-chave: Indeterminação do sujeito; Transmissão Linguística Irregular; Português popular.

ABSTRACT

In this dissertation we study the pronominal strategies expressing the indetermination of the subject in vernacular Portuguese of Bahia, analyzing speech samples from four neighborhoods of the city of Salvador: Liberdade, Cajazeiras, Itapuã and Subúrbio. We present results obtained from the analysis of the social variables whereby the informants are stratified, and interprets them from the perspective of the hypothesis that the current pattern of variation in pronominal strategies of indetermination is the result of changes triggered by the contact of Portuguese with indigenous and African languages during the first centuries of the formation of Brazilian society. We will also observe if there are significant differences between the results found in the different neighbourhoods. The analyses carried out here follow the theoretical and methodological assumptions of Variationist Sociolinguistics (LABOV, 1972) and our hypotheses are directly related to the concept of the Irregular Linguistic Transmission process (BAXTER, 1995; LUCCHESI & BAXTER, 2009) and to the bipolarized reality of Brazilian Portuguese (LUCCHESI, 2015). In our analysis, five pronominal strategies of indeterminacy of the subject were identified: nós, você, eles, a gente and Ø + verb in the third person singular (Ø + V3PS). These variants were analyzed according to their potential conditioning by linguistic variables - phonetic realization of the indeterminate subject, verbal agreement, formal parallelism, reference to the speaker, type of verb, type of sentence, level of agent referentiality and mood of verb; and in terms of the social variables of gender, age, periods spent away from community, schooling and neighbourhood.

Keywords: Indeterminate subject; Irregular Language Transmission; Vernacular Portuguese.

LISTA DE TABELAS

Tabela 1:	População do Brasil por etnia do século XVI ao XIX	22
Tabela 2:	A urbanização do estado da Bahia	30
Tabela 3:	Distribuição dos níveis de abrangência do sujeito indeterminado	33
Tabela 4:	Distribuição das ocorrências de terceira pessoa na fala de Belo Horizonte	43
Tabela 5:	<i>Corpora</i> orais – sujeitos expressos de terceira pessoa	44
Tabela 6:	<i>Corpora</i> escritos – sujeitos expressos de terceira pessoa	44
Tabela 7:	<i>Corpora</i> orais – sujeitos indeterminados	44
Tabela 8:	<i>Corpora</i> escritos – sujeitos indeterminados	45
Tabela 9:	Distribuição dos dados – sentenças infinitivas	46
Tabela 10:	Distribuição das ocorrências de indeterminação do sujeito em cartas do século XIX	46
Tabela 11:	O sujeito nulo em Moçambique	50
Tabela 12:	Distribuição das ocorrências – português e língua nacional	52
Tabela 13:	Enfraquecimento da morfologia verbal no francês	55
Tabela 14:	Enfraquecimento da morfologia verbal no português	56
Tabela 15:	Frequência geral das estratégias de indeterminação do sujeito no português popular de Salvador	79
Tabela 16:	A indeterminação do sujeito de acordo com a variável Realização fonética do sujeito indeterminado	81
Tabela 17:	A indeterminação do sujeito de acordo com a variável Desinência verbal	82
Tabela 18:	A indeterminação do sujeito e a variável Tipo de verbo	83
Tabela 19:	A indeterminação do sujeito de acordo com a variável Nível de referencialidade do sujeito	85
Tabela 20:	Variáveis selecionadas – <i>a gente</i>	87
Tabela 21:	O pronome <i>a gente</i> segundo a variável Nível de referencialidade do agente	88
Tabela 22:	O pronome <i>a gente</i> segundo a variável Realização fonética do sujeito indeterminado	88
Tabela 23:	O pronome <i>a gente</i> segundo a variável Localidade	89
Tabela 24:	O pronome <i>a gente</i> segundo a variável Modo verbal	89
Tabela 25:	O pronome <i>a gente</i> segundo a variável Sexo	90

Tabela 26:	Variáveis selecionadas – <i>eles</i>	91
Tabela 27:	O pronome <i>eles</i> segundo a variável Realização fonética do sujeito indeterminado	91
Tabela 28:	O pronome <i>eles</i> segundo a variável Sexo do informante	91
Tabela 29:	O pronome <i>eles</i> segundo a variável Modo verbal	92
Tabela 30:	O pronome <i>eles</i> segundo a variável Tipo de frase	92
Tabela 31:	O pronome <i>eles</i> segundo a variável Localidade do falante	93
Tabela 32:	Variáveis selecionadas – <i>você</i>	94
Tabela 33:	O pronome <i>você</i> segundo a variável Referência ao falante	94
Tabela 34:	O pronome <i>você</i> segundo a variável Faixa etária	95
Tabela 35:	O pronome <i>você</i> segundo a variável Nível de referencialidade do agente	96
Tabela 36:	O pronome <i>você</i> segundo a variável Realização fonética do sujeito indeterminado	96
Tabela 37:	O pronome <i>você</i> segundo a variável Estada fora da comunidade	97
Tabela 38:	O pronome <i>você</i> segundo a variável Modo	97
Tabela 39:	O pronome <i>você</i> segundo a variável Escolaridade	98
Tabela 40:	O pronome <i>você</i> segundo a variável Localidade	99
Tabela 41:	Variáveis selecionadas - Ø + V3PS	99
Tabela 42:	A estratégia Ø + V3PS segundo a variável Localidade	100
Tabela 43:	A estratégia Ø + V3PS segundo a variável Faixa etária	100
Tabela 44:	A estratégia Ø + V3PS segundo a variável Referência ao falante	101
Tabela 45:	A estratégia Ø + V3PS segundo a variável Estada fora da comunidade	101
Tabela 46:	Variáveis selecionadas – <i>nós</i>	102
Tabela 47:	O pronome <i>nós</i> segundo a variável Sexo	102
Tabela 48:	O pronome <i>nós</i> segundo a variável Localidade	102
Tabela 49:	O pronome <i>nós</i> segundo a variável Faixa etária	103
Tabela 50:	A estratégia <i>nós</i> e a variável Faixa etária no interior da Bahia – extraído de Ponte (2006)	104
Tabela 51:	O pronome <i>nós</i> segundo a variável Realização fonética do sujeito indeterminado	104

Tabela 52:	O pronome <i>nós</i> segundo a variável Nível de referencialidade do Agente	105
Tabela 53:	O pronome <i>nós</i> segundo a variável Tipo de frase	105
Tabela 54:	O pronome <i>nós</i> segundo a variável Escolaridade	106
Tabela 55:	Bairros de Salvador e características socioculturais	109

LISTA DE QUADROS E GRÁFICOS

Gráfico 1:	Distribuição da população rural e urbana do Brasil (1920 – 2010)	29
Quadro 1:	<i>Continuum</i> rural-rurbano-urbano (BORTONI RICARDO, 2006)	31
Gráfico 2:	Sujeitos nulos e plenos	50
Quadro 2:	A estratificação dos informants	66
Quadro 3:	Faixas etárias	77
Gráfico 3:	Estratégias de indeterminação do sujeito no português popular de Salvador	80
Quadro 4:	Paradigma verbal	82
Gráfico 4:	O pronome <i>você</i> e a variável faixa etária	95
Gráfico 5:	Uso da estratégia <i>você</i> em função da escolaridade (CARVALHO, 2008)	98

LISTA DE SIGLAS

NURC: Projeto da Norma Urbana Oral Culta

PEUL: Programa de Estudos sobre o Uso da Língua

TLI: Transmissão Linguística Irregular

SUB: Subúrbio

CAJ: Cajazeiras

ITA: Itapuã

LIB: Liberdade

V1PP: Verbo na primeira pessoa do plural

V3PP: Verbo na terceira pessoa do plural

V3PS: Verbo na terceira pessoa do singular

SUMÁRIO

LISTA DE TABELAS

LISTA DE QUADROS E GRÁFICOS

LISTA DE SIGLAS

INTRODUÇÃO

18

1	O CONTATO ENTRE LÍNGUAS, A TRANSMISSÃO LINGUÍSTICA IRREGULAR E A POLARIDADE DO PORTUGUÊS BRASILEIRO	21
1.1	O <i>CONTINUUM</i> RURBANO E A VARIAÇÃO INTERDIALETAL	28
2	A INDETERMINAÇÃO DO SUJEITO	33
2.1	A IMPESSOALIZAÇÃO E A INDETERMINAÇÃO	34
2.2	A TRADIÇÃO GRAMATICAL	39
2.3	ABORDAGENS VARIACIONISTAS	42
2.3.1	No português brasileiro	42
2.3.2	No português moçambicano	48
2.3.3	No português angolano	51
2.4	O PARÂMETRO DO SUJEITO NULO	53
3	TEORIA E MÉTODO	58
3.1	A ANÁLISE VARIACIONISTA	58
3.1.1	Variação estável e mudança em curso	60
3.2	A CONSTITUIÇÃO DO <i>CORPUS</i>	62
3.2.1	O Projeto Vertentes	62
3.2.2	As comunidades analisadas	62
3.2.2.1	Liberdade	63
3.2.2.2	Cajazeiras	64
3.2.2.3	Itapuã	65
3.2.2.4	Plataforma/Subúrbio	65

3.2.3	O perfil dos informantes	66
3.2.4	As entrevistas	67
3.3	A VARIÁVEL DEPENDENTE	67
3.3.1	A gente	67
3.3.2	Nós/V1PP	68
3.3.3	Você	68
3.3.4	Eles/V3PP	69
3.3.5	Ø + V3PS	69
3.4	AS VARIÁVEIS LINGUÍSTICAS INDEPENDENTES	70
3.4.1	Realização fonética do sujeito indeterminado	70
3.4.2	Desinência verbal	71
3.4.3	Paralelismo formal	71
3.4.4	Referência ao falante	72
3.4.5	Tipo de verbo	72
3.4.6	Tipo de frase	74
3.4.7	Nível de referencialidade do agente	74
3.4.8	Modo	75
3.5	AS VARIÁVEIS SOCIAIS	75
3.5.1	Sexo	76
3.5.2	Faixa etária	76
3.5.3	Estada fora da comunidade	77
3.5.4	Escolaridade	77
3.5.5	Localidade	77
3.6	CONSTITUIÇÃO DOS DADOS E SUPORTE QUANTITATIVO	78
4	ANÁLISE DOS DADOS	79
4.1	DISTRIBUIÇÃO GERAL DA VARIÁVEL DEPENDENTE	79
4.2	NÍVEIS DE REFERENCIALIDADE DO AGENTE	84
4.2.1	Nível genérico	85
4.2.2	Nível parcialmente definido	86
4.2.3	Nível específico	86
4.3	A INDETERMINAÇÃO DO SUJEITO NO PORTUGUÊS	87

4.3.1	A gente	87
4.3.2	Eles	90
4.3.3	Você	93
4.3.4	Ø + V3PS	99
4.3.5	Nós	101
5	CONSIDERAÇÕES FINAIS	107
	REFERÊNCIAS	112
	ANEXOS	118

INTRODUÇÃO

Todas as línguas humanas são dotadas de heterogeneidade, sendo esta o pré-requisito básico para que funcionem plenamente, atendendo às mais diversas necessidades dos seus falantes. Ao contrário do que pregaram por muito tempo os preceitos da linguística moderna, essa heterogeneidade não é aleatória, já que é motivada por fatores internos e externos à estrutura das línguas e passível de descrição e análise. Nesse sentido, propomos, aqui, o estudo da variação dos mecanismos de indeterminação pronominal do sujeito no português popular da cidade de Salvador-Bahia, observando quais são os fatores linguísticos e extralinguísticos que inibem e/ou motivam a variação do fenômeno.

A Gramática Tradicional reconhece como legítimas duas estratégias indeterminadoras na língua portuguesa: aquela em que coloca-se o verbo na terceira pessoa do plural sem pronome expresso, como em (01), e colocando-se o verbo na terceira pessoa do singular com a partícula *se*, como em (02):

(1) *Queimaram* muitos pneus naquele dia.

(2) Fala-*se* muito sobre política na universidade.

Entretanto, muitos estudos acerca do tema na língua falada e escrita (SETTI, 1998; GODOY, 1999; DUARTE, 1993, 1995, 2008; CAMPOS, 2010; ASSUNÇÃO, 2010, entre outros) mostram que a língua portuguesa apresenta muitas outras estratégias de indeterminação, que incluem o uso dos pronomes *a gente*, *nós*, *você*, *eles*, bem como estratégias sem sujeito expresso, como o uso da primeira pessoa do plural e da terceira pessoa do singular sem o pronome *se*:

(3) *A gente* tem que ter fé em Deus.

(4) *Nós* brincávamos muito naquele rio.

(5) Antigamente, *você* dormia de porta aberta aqui.

(6) *Eles* vieram aqui à noite e *mataram* os meninos.

(7) *Devemos* acreditar que as coisas vão melhorar.

(8) Ø Disse que aqui é um bom lugar para viver.

Estratégias pronominais como as exemplificadas acima serão o centro do estudo proposto aqui, que será realizado no âmbito da Sociolinguística Variacionista e terá como *corpus* 48 entrevistas recolhidas em quatro bairros da cidade de Salvador: Liberdade, Subúrbio, Itapuã e Cajazeiras. O foco da análise é o português popular da Bahia (doravante PP), ou seja, a variante utilizada por pessoas com pouca ou nenhuma escolaridade, pois acreditamos que o contato entre línguas nas circunstâncias de formação da sociedade brasileira foi crucial para a formação das características atuais da variante popular do português brasileiro. Para tanto, pretendemos:

1. Descrever as formas pronominais utilizadas como estratégia para indeterminar o sujeito em Salvador, pelos falantes com pouca ou nenhuma escolaridade;
2. Verificar os condicionamentos linguísticos e sociais que motivam a escolha de uma ou outra estratégia.

A nossa principal hipótese é que as circunstâncias sócio-históricas que envolveram a formação do português brasileiro deixaram marcas que podem ser retomadas a partir da análise do português popular, sendo que essas marcas estariam mais latentes na fala dos informantes mais velhos e/ou que não foram escolarizados. Temos como hipótese, também, que a variante da terceira pessoa do singular sem o pronome *se* é fruto da não aquisição desse pronome durante a formação do PB, caracterizando as não incorporações morfológicas típicas das línguas formadas em situações de contato. Para que se confirme ou refute essas hipóteses, analisaremos as entrevistas, identificando todas as ocorrências de sujeito indeterminado. Em seguida, as hipóteses sobre o condicionamento linguístico e sociolinguístico serão avaliadas quantitativamente, pelos procedimentos estatísticos do programa GOLDVARB X (cf. Capítulo de Metodologia).

Para o desenvolvimento argumentativo proposto aqui, essa dissertação será estruturada da seguinte forma:

No primeiro capítulo, faremos uma discussão das circunstâncias sócio-históricas que balizaram a formação do português brasileiro, discutindo, também, as diferentes teorias que explicam a formação da variante brasileira do português e abordando a hipótese da polaridade do português brasileiro, o conceito de Transmissão Linguística Irregular e o *continuum* rural-urbano.

No segundo capítulo, discutiremos o conceito de indeterminação, abordando o que é proposto pelas gramáticas tradicionais. Apresentaremos algumas abordagens variacionistas do tema, discutindo os resultados encontrados pelos pesquisadores. Falaremos, também, sobre pesquisas feitas em outras variedades do português, como a moçambicana e a angolana, e sobre as diferenças e semelhanças entre os conceitos de indeterminação e impessoalização.

No capítulo três, discutiremos a relação entre os sujeitos de referência definida e os de referência arbitrária, trazendo resultados de estudos desse parâmetro e mostrando como ele pode afetar os sujeitos indeterminados.

No quarto capítulo apresentamos a fundamentação teórica e metodológica do nosso estudo: a Sociolinguística Variacionista, bem como detalharemos a amostra utilizada, traçaremos um perfil dos informantes e do *corpus* e descreveremos a variável dependente, as variantes que a compõem, as variáveis linguísticas e sociais.

No capítulo cinco, apresentaremos os resultados obtidos e faremos considerações acerca das variáveis linguísticas e sociais que foram consideradas estatisticamente relevantes para cada uma das variantes.

No sexto e último capítulo, Considerações finais, sintetizaremos os resultados obtidos, sinalizando os pontos mais relevantes e apontando possíveis direcionamentos futuros da pesquisa.

1 O CONTATO ENTRE LÍNGUAS, A TRANSMISSÃO LINGUÍSTICA IRREGULAR E A POLARIDADE DO PORTUGUÊS BRASILEIRO

Neste capítulo, comentaremos brevemente os parâmetros sócio-históricos que motivaram a análise realizada aqui, mencionando algumas teorias acerca da formação do português brasileiro. Discutiremos o conceito de Transmissão Linguística Irregular e comentaremos, também, as consequências do rápido e intenso processo de urbanização e industrialização pelo qual o país passou no decorrer do século XX, no intuito de compreender o efeito do contato interdialetoal nas grandes cidades, como é o caso de Salvador.

Os estudos acerca do funcionamento das línguas que se formaram através de processos intensos de contato entre línguas tipologicamente diferentes vêm, nos últimos anos, se tornando cada vez mais frequentes no âmbito da análise sociolinguística no Brasil, sendo que tais estudos focalizam, além do plano formal de estudo, as razões sócio-históricas responsáveis pelas mudanças sofridas por essas línguas. É por esse caminho que autores como Lucchesi (1994, 1996, 2001, 2002, 2003, 2006, 2015), Baxter (1998, 2003, 2006), Baxter e Lucchesi (1993, 1997, 2006), Castro (2002), Holm (1987, 1988, 1992, 2004), Mattos e Silva (1988, 2001, 2002, 2004), dentre muitos outros, encaminham os seus estudos sobre as origens do português brasileiro.

Em contraste à falsa ideia de que o Brasil é um país monolíngue, é estimado que haja, atualmente, cerca de 180 línguas indígenas, que são faladas por cerca de 220.000 indígenas que vivem no Brasil, números que mostram o glotocídio pelo qual essas línguas passaram no período colonial (estima-se que existiam por volta de 1200 línguas antes da chegada dos portugueses, o que significa uma diminuição de 85% do total de línguas que existiam no território brasileiro) (MATTOS E SILVA, 1988, p. 17). Os primeiros portugueses, ao chegar ao território do Brasil, adotaram como instrumento de comunicação e de subjugação as línguas gerais, termo que designa um grupo de línguas que surgiram a partir do contato entre os índios tupi-guaranis e os europeus: a língua geral da costa (ou língua geral basílica), de feições próximas à língua tupi, a língua geral cariri, na região nordeste, e na língua geral amazônica, de base tupinambá, que ainda é falada.

Esse quadro, entretanto, começa a se modificar com o progresso do empreendimento agricultor da colônia. Segundo Lucchesi e Baxter (2009, p.44-45)

a resistência cultural intrínseca do índio ao trabalho forçado, sobretudo ao trabalho agrícola (que na sua cultura nômade extrativista ocupava uma posição subalterna, sendo entregue às mulheres e crianças), aliada às campanhas contra a escravidão indígena movidas pelos jesuítas, fez com que se fizesse necessário buscar uma outra fonte de mão de obra para atender as

crescentes demandas dos emergentes e muito prósperos engenhos de cana de açúcar, que começaram a se instalar no Nordeste brasileiro, já no século XVI, e se converteram muito rapidamente no setor econômico mais dinâmico do empreendimento colonial do Brasil. Assim, o sequestro e transporte de populações africanas — que passou à história com a denominação de tráfico negreiro — forneceu a força de trabalho para a implementação da cultura agroexportadora do açúcar, que viveria o seu apogeu durante o século XVII. E, certamente, os fabulosos lucros proporcionados pelo tráfico de escravos africanos para servirem aos senhores de engenhos, ávidos por mão de obra para impulsionar o seu próspero empreendimento, falou mais alto do que qualquer resistência cultural indígena ou campanha humanitária de bem intencionados missionários.

A partir de então, visando aos lucros proporcionados pelas lavouras de açúcar e considerando o insucesso da tentativa de escravizar os indígenas, que encontrava apoio nas campanhas jesuíticas, inicia-se o tráfico de africanos para o Brasil, mais precisamente no ano de 1538, embora a autorização só tenha sido dada em 1549, através de um alvará assinado por D. João III, autorizando o tráfico de até 120 “peças” por engenho de açúcar, segundo Raimundo (1933, p.2627, apud LUCCHESI, 2009, p. 45).

O peso demográfico do contingente africano para a formação da sociedade brasileira e da língua falada no Brasil é indiscutível. Fala-se que, durante os mais de 300 anos de escravidão, foram trazidos para o Brasil entre 3.500.000 e 3.600.000 africanos, sendo que, em termos demográficos, essa população chegou a ultrapassar a proporção de brancos no território brasileiro, conforme pode-se observar no quadro a seguir:

Tabela 1 - População do Brasil por etnia do século XVI ao XIX

ETNIA	1583-1600	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%
Branco brasileiros	-	5%	10%	17%	24%
Europeus	30%	25%	22%	14%	17%
Índios integrados	50%	10%	8%	4%	2%

Fonte: Mussa, 1991, p.163

Esses povos eram sequestrados da área oeste-africana e da área banta, onde localizam-se, hoje, o Congo, República do Congo e Angola, e a Costa Oriental (Moçambique) (PETTER, 2006, p.124-125), sendo, em terras brasileiras, submetidos às mais variadas

situações de privação de direitos. Assim, a aquisição e propagação da língua portuguesa por parte dos africanos escravizados acontecia paralelamente ao movimento dos ciclos econômicos do Brasil: primeiro os engenhos de açúcar e tabaco, depois o ciclo do ouro em Minas gerais, depois as lavouras de café. Sobre o estatuto das línguas trazidas pelos africanos para o Brasil, Petter (2006, p.126) afirma que

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

A partir de então, importantes acontecimentos sócio-históricos definiram o rumo do português brasileiro. O significativo aumento da população branca e mestiça com o ciclo do ouro, e, posteriormente, com as lavouras cafeeiras, o surgimento de importantes centros urbanos, como Vila Rica e São João Del Rey, os efeitos da mudança da capital do Brasil para o Rio de Janeiro, o aumento vertiginoso da população negra e mestiça e o impacto demográfico da fuga da família real para o Brasil foram significativos para definir as características da língua que se formava (LUCCHESI, 2009, p.49). Entretanto, é possível que se afirme, com certa margem de segurança, que nenhum acontecimento teve maior impacto sociolinguístico que a abolição da escravatura, em 1888 e, posteriormente, a industrialização, que possibilitou a urbanização intensa do país. Com o fim da escravidão, muitos ex-cativos ficaram em situações sociais de miséria. Alguns se voltaram para a cultura de subsistência, se dirigindo para o interior. Confinados nessas áreas, eles puderam, então, conservar e difundir o português com as mudanças características das línguas adquiridas em situação de contato, enquanto os centros urbanos difundiam a língua sob influência direta dos modelos portugueses.

É quase unânime entre os estudiosos da área a afirmação de que as línguas indígenas pouco influenciaram a estrutura da língua que se formava, já que os ameríndios, ao serem vertiginosamente exterminados, não atuaram de forma determinante na estrutura gramatical

do português brasileiro. Assim sendo, e considerando as evidências apontadas anteriormente, torna-se viável assumir que os africanos e afrodescendentes foram os principais responsáveis pela formação e difusão da variante popular do português, ficando a cargo da elite europeia a propagação da variante culta da do português brasileiro, embora a importância do contato massivo para a formação do português popular não seja admitida por todas as hipóteses explicativas da formação do português brasileiro.

As hipóteses explicativas para a formação da variante popular são múltiplas. A primeira hipótese, formulada pelo linguista norte americano Gregory Guy (1981), relaciona, sobretudo através de evidências linguísticas e sócio-históricas comparadas, as características do português brasileiro a um processo de criouliização prévia, sendo que esse crioulo português teria passado por um rápido processo de descriouliização. Apesar de suas hipóteses serem balizadas fundamentalmente por razões linguísticas, haja vista que Guy traça paralelos entre construções do português popular brasileiro, tais como como concordância nominal e verbal, associando-as a modelos encontrados em línguas crioulas consolidadas, o autor não desconsidera a importância dos fatores históricos e sociodemográficos que contribuíram para as características do português brasileiro.

A segunda hipótese de origem do português popular foi formulada por Antony Naro e Marta Scherre e descrita em vários artigos da sua autoria desde a década de 1990. Para esses linguistas, o contato massivo entre línguas teve um papel pouco relevante, sendo que as características do português brasileiro são fruto de tendências internas, imanentes à língua e previstas para todas as línguas do tronco indo-europeu e, conseqüentemente, das línguas românicas em direção à uma simplificação morfológica, processo denominado Deriva Secular. Entretanto, os fatores empíricos que contribuíram para que essas mudanças internas acontecessem, chamados pelos autores de “confluência de motivos”, são pouco claros (LUCCHESI, 2012, p.268):

[...] até hoje, os fatos que Naro e Scherre conseguiram sistematizar não têm sido reconhecidos como evidências incontestes contra o papel que o contato entre línguas teria desempenhado na formação do português do Brasil. Os dados do português arcaico têm recebido questionamentos de natureza filológica e os dados das variedades rurais e populares do português europeu contemporâneo vão de encontro à opinião corrente na comunidade de linguistas portuguesas.

Conforme mencionamos anteriormente, o início do tráfico negreiro, que trouxe ao território brasileiro um contingente gigantesco de africanos, fez com que as características do português popular brasileiro começassem a ser definidas mais precisamente. Devido às suas

necessidades de rápida adaptação e sobrevivência, os africanos sequestrados precisaram adquirir rapidamente o português, não contando, para essa aquisição, com a escolarização. Da mesma forma se dava a transmissão da língua para os descendentes dessa população, já que essas crianças tinham como modelo de língua materna uma versão alterada de português, falado como segunda língua. Essas peculiaridades sócio-históricas dos processos de aquisição da língua por parte das populações africanas e mestiças são o que nos permite afirmar que o português brasileiro se formou através de um processo de Transmissão Linguística Irregular¹ (LUCCHESI, 2003) (doravante TLI). A variante da língua que se formou, com as suas alterações e simplificações morfológicas, se mantém até os dias de hoje, na fala das camadas economicamente desfavorecidas (LUCCHESI et al, 2009, p.28):

Apesar de reunir condições sócio-históricas, em princípio, muito propícias à criouliização da língua do colonizador europeu, não ocorreu no Brasil um processo estável, duradouro e representativo de criouliização da língua portuguesa. Mas, se a criouliização do português no Brasil foi, na melhor das hipóteses, um fenômeno historicamente efêmero e localizado, não se pode pensar seriamente que a língua portuguesa não foi diretamente afetada pelo contato do português com as línguas africanas de uma forma bem ampla e representativa, até porque os afrodescendentes se integraram em todos os segmentos sociais e nos mais diferentes ramos da atividade econômica, em todas as regiões do país; concentrando-se, porém, na base da pirâmide social, em função das adversidades históricas que tiveram de enfrentar.

Para os parâmetros epistemológicos da TLI, se um falante adulto aprende uma língua nas circunstâncias como as observadas na formação do português brasileiro – de forma emergencial, sem o intermédio homogeneizador da escola e, conseqüentemente, com dados irregulares relativos à estrutura da língua-alvo – ele fornece para os descendentes um modelo de segunda língua que traz uma forte ausência/variação de estruturas gramaticais, sendo que apenas os elementos mais frequentes e salientes, básicos para a realização da comunicação, são mantidos, sendo que Os descendentes nativizam a língua com todas essas características de aquisição e repetem para os seus sucessores o mesmo processo. No caso do português brasileiro, observa-se que a língua que foi nativizada não sofreu as reestruturações gramaticais radicais que caracterizam as línguas crioulas, e sim pela não-incorporação de algumas estruturas gramaticais, o que vai ao encontro da hipótese de que o português brasileiro emergiu de um processo de TLI não radical.

¹ Encontramos, em Winford (2000, 2003) e Holm (2000), perspectivas semelhantes, com evidências da relevância dos processos de referência e incorporação de material de L1 sobre a L2 em formação, bem como a generalidade das perdas morfológicas que também ocorrem em processos de criouliização.

É importante se considerar, também, os fatores sócio-históricos que fizeram com que, no Brasil, não se consolidasse uma língua crioula estável. Sendo assim, os fatores a seguir são apontados como os motivos extralinguísticos que fundamentam a hipótese de que o português brasileiro emergiu não de um processo radical de Transmissão Linguística Irregular, que teria resultado em um produto mais típico de crioulação, mas de um processo de Transmissão Linguística Irregular de tipo leve, segundo Baxter e Lucchesi (2006, p. 189):

- a) a proporção entre a população de origem africana e branca, que permitia um nível de acesso maior à língua-alvo do que o observado nas situações típicas de contato, considerando que o nível de crioulação é inversamente proporcional ao acesso dos falantes à língua-alvo;
- b) a ausência de vida social e familiar entre as populações de escravos, provocada pelas condições sub-humanas de exploração, pela alta taxa de mortalidade e pelos sucessivos deslocamentos, o que impedia ou dificultava muito a nativização de uma possível língua crioula, sendo que essa socialização é indispensável para a emergência da língua;
- c) o uso das línguas francas africanas como instrumento de interação dos escravos segregados e foragidos: ao invés de se comunicar através de uma possível língua crioula incipiente, os africanos escravizados privilegiavam o uso de línguas francas baseadas suas línguas nativas;
- d) o incentivo à proficiência do português;
- e) a maior integração social dos escravos urbanos, domésticos e das zonas mineradoras;
- f) a miscigenação racial.

Esses fatores sócio-históricos e estruturais são o que nos permite afirmar que o português popular brasileiro, em linhas gerais, foi formado através de um processo de TLI leve, sendo que, se houve, no território nacional, processos de transmissão irregular mais radicais, no sentido de uma crioulação mais típica, tais processos ficaram restritos a regiões isoladas e foram rapidamente absorvidos. Como herança, temos uma língua plural, heterogênea, e polarizada em dois grandes padrões coletivos: a norma culta e a norma popular (LUCCHESI, 2006, 2009, 2015). Descreveremos essas normas a seguir.

A norma culta, variante da língua portuguesa defendida pela elite conservadora, cuja modalidade linguística deriva da variante falada pela elite europeia do Brasil colonial, é, de acordo com os critérios de classificação do NURC, a fala da população favorecida economicamente, oriunda dos centros urbanos e detentora de um ideal de perfeição de uso da língua, sendo que seu uso é favorecido por contextos de fala mais monitorados. Os falantes dessa variante conservam os traços de fala que os distinguem das classes consideradas inferiores durante o processo de formação do português brasileiro.

A modalidade popular, com origem atribuída por diversos autores, tais como Silva Neto, Lucchesi, Baxter, entre outros, ao modelo de língua nativizado pela população menos abastada, descendente dos africanos escravizados, dos indígenas e da população mulata e mestiça durante o processo de contato entre línguas e formação do português brasileiro, é utilizada, atualmente, pelas classes menos favorecidas socialmente, habitantes, geralmente, das áreas menos urbanizadas, é aquela que goza de menor prestígio social. Se, nos primeiros séculos de formação da sociedade brasileira, essa variante carregava, de forma intensa, as evidências do processo de contato entre línguas, ela perdeu, sobretudo nas grandes cidades, algumas dessas evidências, devido à forte e inegável influência, principalmente, dos meios de comunicação e da urbanização. Ainda assim, o português popular ainda é alvo de estigma por ser a fala de pessoas com pouca ou nenhuma escolaridade, sendo, muitas vezes, considerada como errada ou inadequada, devido aos seus desvios morfossintáticos, lexicais e fonológicos em relação à norma que é socialmente aceita.

No caso da indeterminação do sujeito, a hipótese é a de que o contato massivo entre línguas seria responsável pelo atual panorama do fenômeno: temos um quadro com ampla variação entre os recursos de indeterminação, bem como uma variante que pode ter sido resultado da não aquisição do mecanismo gramatical de indeterminação com o pronome *se*: a terceira pessoa do singular sem referência anterior. No decorrer da análise, buscaremos testar essa hipótese.

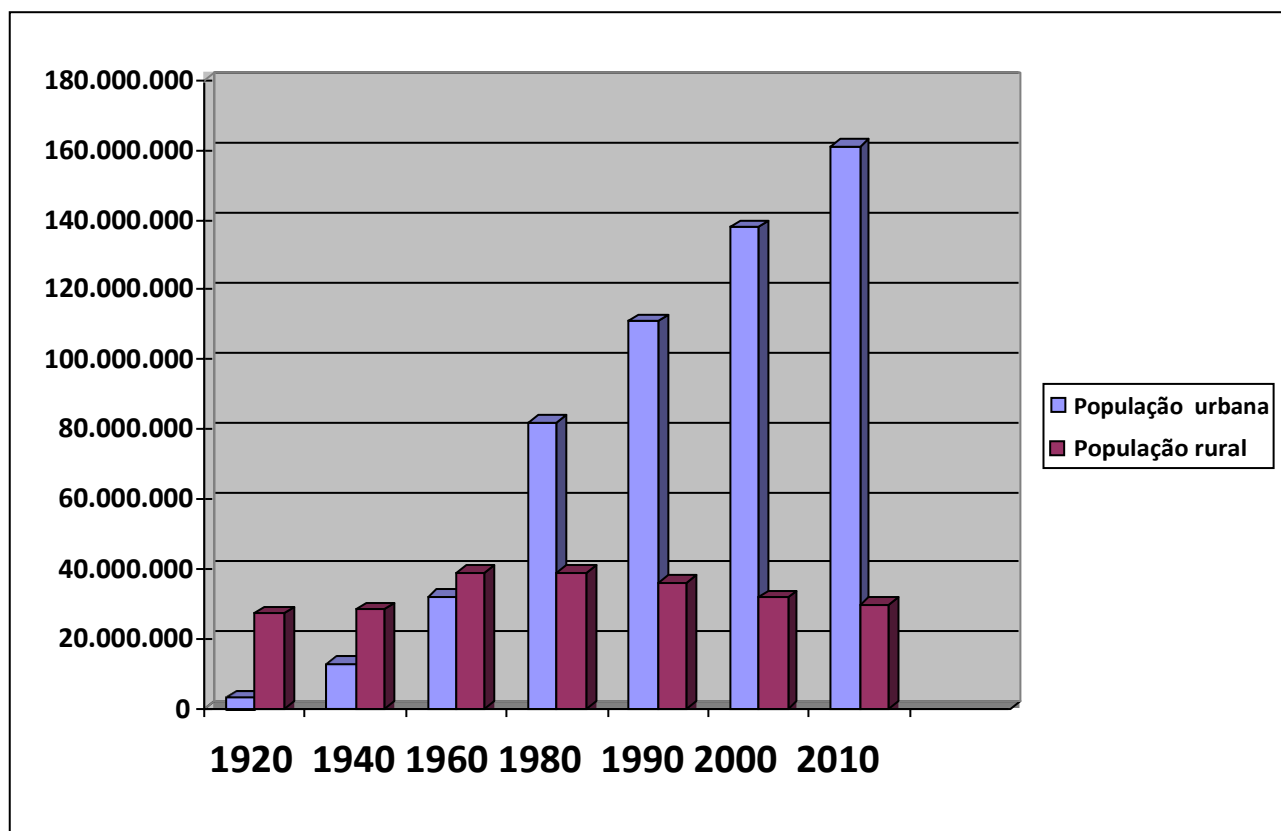
1.1 O *CONTINUUM* RURBANO E A VARIAÇÃO INTERDIALETAL

Nesse ponto, é válido falar sobre o processo de urbanização do Brasil em meados do século XX, já que esse o mesmo ocasionou importantes implicações sociolinguísticas, sobretudo no que diz respeito à dinâmica do contato interdialetoal nas grandes cidades, como é o caso de Salvador. A industrialização e urbanização possibilitaram o acesso da população rural, até então majoritária no Brasil, aos grandes centros urbanos, através de estradas que ligavam o interior à cidade grande, e acesso, por parte dessa população, aos meios de comunicação, diminuindo a distância entre o interior e a capital (LUCCHESI et al, 2009, p.54-55):

No bojo desse processo, são construídas as rodovias que integram todas as regiões do território nacional, e os meios de comunicação de massa experimentam um crescimento vertiginoso, alcançando e influenciando — inicialmente, através do rádio e, depois, através da televisão — até as comunidades mais isoladas dos mais afastados rincões do território brasileiro. Esse gigantesco processo, que alterou profundamente a fisionomia do país (mesmo sem o livrar de suas deformações estruturais), acabou por atenuar, em sua ação niveladora, a acentuada polarização entre a fala das elites urbanas e a fala das populações rurais pobres; tanto ao eliminar, na fala popular, as marcas mais características do processo de transmissão linguística irregular, ocorrido nos séculos anteriores (ou mesmo provocando o desaparecimento, em algumas localidades, de formas incipientes de línguas crioulas); quanto ao permitir que certas mudanças ocorridas na fala popular penetrassem na fala das camadas médias e altas. Nesse caso, o grande fenômeno da imigração ocorrido entre os finais do século XIX e as primeiras décadas do século XX pode ter desempenhado um papel decisivo.

Lucchesi (2015) aponta como marco inicial para a interpretação das mudanças ocasionadas pelos processos de urbanização e industrialização a revolução de 1930, ano que marca o fim do período histórico brasileiro denominado República Velha, sendo o governo subsequente, chefiado pelo presidente Getúlio Vargas, marcado principalmente pelo incentivo à urbanização e industrialização no país. A partir desse período, a população, que, até então, vivia isolada no interior do país, passou, motivada pelo crescimento das atividades urbanas, a ocupar os grandes centros em um intenso movimento de êxodo rural, que fica mais bem evidenciado no gráfico a seguir:

Gráfico 1- Distribuição da população rural e urbana do Brasil (1920-2010)



Fonte: LUCCHESI (2015, p.144) adaptado

O gráfico nos mostra que a estimativa da população urbana brasileira, entre os anos de 1920 e 2010, aumentou em mais de 150 milhões de habitantes, enquanto a população rural manteve-se praticamente inalterada nesse período. Com isso, as variedades linguísticas do interior do país, antes isoladas, passam a ocupar o perímetro urbano, sendo que as diferenças, até então diatópicas, passam a ser diastráticas e, com o tempo, passaram a ser dirimidas pelo contato com o dialeto culto, com a escolarização e com os meios de comunicação em massa (LUCCHESI, 2015).

O caso específico da urbanização do estado da Bahia e da reestruturação do espaço urbano da cidade de Salvador se configurou em moldes semelhantes aos do resto do país, mas é importante considerar que, na Bahia, a implementação da industrialização não priorizou a cidade de Salvador, mas sim algumas cidades em seu entorno, como São Francisco do Conde e Camaçari, principalmente após a década de 1960 (SANTOS, 2013). De acordo com a análise realizada por Silva (1988), pode-se resumir no quadro a seguir os fatores que motivaram a urbanização no estado da Bahia, tanto em relação ao estado como um todo quanto no que diz respeito à região metropolitana, apontando as principais mudanças

econômicas que motivaram a mudança na dinâmica social da Bahia, que tiveram importantes implicações no cenário sociolinguístico do estado:

Tabela 2 - A urbanização do estado da Bahia

Planos analíticos Dimensões espaciais	Plano externo	Plano interno
Nível estadual	<p>Mudanças na estrutura e no relacionamento da economia nacional em um contexto internacional.</p> <p>Mudanças no sistema de transporte a nível nacional, em termos de estrutura em rede.</p> <p>Mudanças no papel exercido pelas cidades, em termos sistêmicos.</p>	<p>Mudança no relacionamento da economia estadual e de sua estrutura econômica.</p> <p>Expansão e retração de economias sub-regionais e urbanas.</p> <p>Mudança nas funções urbanas.</p>
Nível metropolitano	<p>Necessidade de uma desconcentração da economia nacional, exigindo um vigoroso papel do setor público em determinadas áreas.</p> <p>Necessidade de integração da economia nacional, exigindo uma adequação do sistema de transportes e comunicações em regiões específicas.</p>	<p>Dinamismo das atividades econômicas metropolitanas, sobretudo industriais.</p> <p>Reorganização de áreas urbanas e implantação de novos eixos de expansão metropolitana.</p>

Fonte: SILVA (1988, p.32)

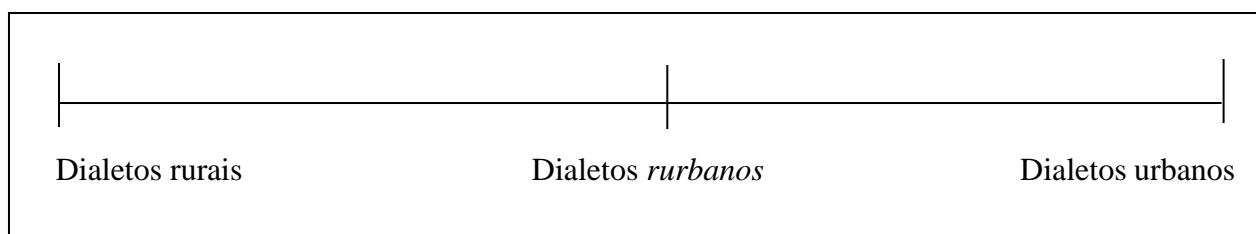
Todas essas modificações pelas quais o Brasil, de modo geral, e o estado da Bahia, em particular, passaram, provocaram modificações no contexto social e econômico do país. Entretanto, os grandes centros não conseguiram absorver toda a população que chegava às cidades em busca de inserção no mercado de trabalho, o que fazia com que essas pessoas se aglomerassem na periferia dessas cidades.

Sendo assim, essa configuração social pós urbanização inviabiliza a afirmação de que a realidade linguística brasileira está dividida em uma dicotomia rígida entre as normas cultas e populares, sobretudo quando analisamos a diversidade linguística dos grandes centros, como Salvador. Nesses falantes, como afirma Cyranda et al (2010, p. 365), “se reconhecem os chamados os chamados traços graduais, que se caracterizam por terem uma distribuição descontínua nas áreas urbanas. Poderíamos neles reconhecer mesmo uma gradação, isto é, a presença desde construções muito estigmatizadas até outras consideradas padrão”.

Considerando, então, a aproximação ocorrida entre campo e cidade, é válido comentar um modelo bastante difundido atualmente para representar a diversidade do português brasileiro: o modelo de Bortoni-Ricardo (2006). Levando em consideração o rápido processo de urbanização do país, a autora descreve a realidade linguística brasileira como contida em três *continua*: o *continuum* rural-urbano, o oralidade-letramento e o da monitoração estilística, tendo em vista que os falantes apresentam mais de um estilo de fala, que dependerá das atividades comunicativas e das práticas sociais nas quais os mesmos estão inseridos.

Em relação ao contexto rural-urbano, a autora afirma que as variedades do português brasileiro estão distribuídas em um esquema de *continuum*, sendo que a diferenciação entre essas variedades se dá de acordo com o grau de isolamento e com o prestígio social de cada uma delas, estando os dialetos *rurbanos* na posição intermediária desse *continuum*:

Quadro 1- *Continuum* rural-rurbano-urbano (BORTONI RICARDO, 2006)



Assim sendo, a fala urbana comum teria muitas características em comum com a norma culta, o que caracteriza os dialetos *rurbanos*. Sobre esse modelo, Faraco (2008, p.44) afirma que

Considerando as características da urbanização no país (que, em menos de cinquenta anos, inverteu a distribuição da população entre o campo e a cidade, tornando o Brasil um dos países mais urbanizados do mundo, com aproximadamente 80% de sua população vivendo hoje nas cidades), e o alcance dos seus meios de comunicação social (o rádio está praticamente em todos os lares brasileiros e a televisão, com produção e transmissão fortemente centralizadas, chega a mais de 90% deles), podemos dizer que as variedades que exercem, hoje, a maior força de atração sobre as demais são as faladas pelas populações tradicionalmente urbanas, situadas na escala de renda de média para alta e que, por isso, têm garantido para si, historicamente bons níveis de escolaridade (pelo menos a educação média completa) e o acesso aos bens de cultura escrita.

Afirmamos, então, que, dada a complexidade do cenário linguístico brasileiro, a correlação com a sócio-história do país é fundamental para compreender os fenômenos que se encontram em variação na língua, quadro no qual se incluem os recursos de indeterminação do sujeito. É através das vicissitudes de formação do PB e da compreensão dos acontecimentos históricos que motivaram a grande transformação social pela qual o país passou no decorrer do século XX que podemos compreender os motivos que levam à variação e às perdas morfológicas pelas quais o fenômeno da indeterminação vem passando.

Considerar os processos que balizaram a formação do português brasileiro e os aspectos históricos e sociais do país depois disso nos leva a perceber o quanto é complexa a realidade linguística do Brasil e, grosso modo, de todas as demais variedades do português que se formaram através de processos de contato massivo entre línguas. Portanto, além de pensar nas características do português brasileiro em comparação às do português europeu, é importante, para que possamos compreender mais a fundo os efeitos do contato entre línguas, a análise de outras variedades do português formadas nessas circunstâncias, como é o caso do português angolano e do português moçambicano, por exemplo, conforme falaremos posteriormente.

Neste capítulo, traçamos um panorama sócio-histórico das circunstâncias de formação do português brasileiro, considerando a importância das línguas indígenas e africanas para a formação do português popular brasileiro. Estudar relação entre as normas cultas e populares e a dinâmica do contato interdialeto, bem como o processo de urbanização e industrialização do país, é fundamental para que possamos compreender a configuração do atual português popular, sobretudo nas grandes cidades, como é o caso de Salvador, e para explicar as modificações pelas quais os fenômenos morfossintáticos, como é o caso dos recursos de indeterminação do sujeito, passaram.

2 A INDETERMINAÇÃO DO SUJEITO

A indeterminação do sujeito é um fenômeno linguístico que motiva uma vasta quantidade de estudos no português brasileiro. Vários autores debruçam-se sobre as diversas possibilidades pronominais, lexicais e pragmáticas de se indeterminar o sujeito, pautando-se em *corpus* do *continuum* oral-escrito/culto-popular e considerando a possibilidade de se inserir estratégias inovadoras.

Milanez, em sua dissertação, defendida em 1982, foi a primeira autora a ressaltar a importância de se estudar a indeterminação do sujeito em um nível semântico, afirmando que o fenômeno não se apresenta de forma absoluta, mas sim diferenciando-se em graus de abrangência que envolvem todas as pessoas do discurso. A autora distribui esses níveis da seguinte forma (MILANEZ, 1982, p. 80, 81):

Tabela 3 - Distribuição dos níveis de abrangência do sujeito indeterminado

Níveis mais abrangentes	
Ø + terceira pessoa do singular, Ø + terceira pessoa do plural + <i>se</i> , Ø + infinitivo	Devido à falta de marcação de pessoa, distanciam-se das referências definidas.
Níveis intermediários	
<i>A gente, você, eu, nós</i>	São menos genéricos, por ainda manter referências específicas aos protagonistas da situação dialogal.
Níveis menos abrangentes	
Ø + terceira pessoa do plural, <i>Eles</i>	Formas restritas ao universo das terceiras pessoas.

Sobre essa classificação, a autora faz importantes constatações:

- Apesar de se localizarem no mesmo nível, as estratégias *a gente* e *nós* não se comportam da mesma forma que as estratégias *você* e *eu*, já que aquelas incluem o falante de forma real, e estas envolvem os falantes de forma hipotética;

- O pronome *eles* e a estratégia \emptyset + terceira pessoa do plural podem levantar dúvidas em relação à sua classificação, podendo ser vistos como ocupando o limite entre a indeterminação e a indefinição. Para a autora, apesar de aparentemente restritas ao universo das terceiras pessoas, essas estratégias apresentam efeitos generalizadores ilimitados, o que nos leva a concluir que, além da forma, o contexto, que pode determinar a abrangência de forma mais eficaz do que a própria forma. Assim sendo, também consideraremos, na análise aqui apresentada, essas duas formas como fazendo parte do *continuum* da indeterminação.

2.1 A IMPESSOALIZAÇÃO E A INDETERMINAÇÃO

Nessa seção, trataremos de uma importante característica sintático-semântica das línguas humanas e da sua relação com os mecanismos de indeterminação do sujeito: a impessoalização. A descrição e análise dos pronomes impessoais nas línguas humanas se constituem em um desafio para a linguística, uma vez que as propriedades sintáticas e semânticas desses pronomes não são facilmente generalizáveis. Entretanto, a análise da impessoalização pode prestar importantes esclarecimentos acerca da forma como os falantes contextualizam o conteúdo semântico dos sujeitos e a falta de agentividade do argumento externo.

A seguir, iremos discutir algumas propostas de tipologia de construções impessoais, com vistas a analisar as diferentes formas de tratar a impessoalidade.

Podemos definir a impessoalização a partir do próprio conceito de sujeito. Se considerarmos que a existência do sujeito oracional pressupõe a existência de agentividade, podemos definir a impessoalidade como a ausência ou como a não especificação dessa agentividade (SIEWIERSKA, 2007). Kitagawa & Lehrer (1990), na busca de uma definição do conceito, afirmam que o antecedente dos pronomes impessoais pode ser “anyone”, no caso da ausência de agentividade, ou “everyone”, no caso da não especificidade. Porém, alguns autores afirmam que a heterogeneidade das construções impessoais não permite que haja uma definição única. Creissels (2008), por exemplo, através de uma análise morfossintática dos pronomes impessoais, afirma:

Given the lack of consensus in the use of the term impersonal, and the obvious heterogeneity of the constructions to which it has been applied in various traditions, I will not try to propose a general definition of these constructions. I will rather propose to deal separately with at least three types of constructions I will call respectively simple, special and covert impersonal constructions.²

Sendo assim, o autor trata essas três construções da seguinte forma:

a) Construções impessoais simples

De acordo com o autor, essas construções ocorrem com verbos tal qual em construções canônicas, sendo que não há alteração na estrutura dos termos que não são sujeito. Nessas construções, a interpretação arbitrária do sujeito se dá através de outros meios, e não por um sintagma nominal canônico indefinido no papel de sujeito. Há, na análise do autor, duas possibilidades de classificação dessas construções: as não marcadas, com um sintagma nominal referencial no lugar de sujeito, e marcadas, apenas com marcação morfológica verbo.

b) Construções impessoais especiais

Essas construções envolvem duas possibilidades: as construções especiais pragmáticas, com verbos cujo argumento externo tem características de paciente, e as construções impessoais léxico-semânticas, que podem constituir a única construção possível do verbo que envolvem ou necessitar de uma estrutura argumental diferente da vista em construções canônicas que envolvem o mesmo verbo.

c) Construções impessoais encobertas

São construções que incluem um termo com características de sujeitos canônicos, mas não apresentam algumas propriedades comportamentais correspondentes, como a possibilidade de receber papel temático. (SOUZA, 2013, p. 64)

Em sua avaliação da tipologia dos pronomes pessoais, Gast (2015) aplicou as seguintes propriedades de análise e interpretação dos eventos que envolvem pronomes impessoais, com o intuito de traçar generalizações acerca dos tipos de contextos em que os pronomes impessoais podem ocorrer, analisando um banco de dados de 188 estratégias em 49 idiomas (GAST, 2015, p.6):

1. Veridicalidade, que avalia se a sentença tem ou não valor verídico;

a) Com valor verídico: *One only lives once (Só se vive uma vez).*

²“Dada a falta de consenso no uso do termo “impessoal”, e a óbvia heterogeneidade das construções às quais foi aplicado em várias tradições, não tentarei propor uma definição geral para essas construções. Prefiro tratar separadamente com pelo menos três tipos de construções que eu chamarei, respectivamente, de construções impessoais simples, especiais e encobertas.” (Tradução nossa)

- b) Com valor não verídico: One shouldn't drink and drive (Não se deve beber e dirigir).
2. A quantificação do evento expresso pelo contexto onde se encontra o pronome, podendo esta ser esporádica ou genérica;
- a) Genérico: One only lives once (Só se vive uma vez).
 - b) Esporádico: One could see that he was drunk (Podia se ver que ele estava bêbado).
3. Presença ou ausência de verbos modais;
- a) Presença: One could see that he was drunk (One could see that he was drunk).
 - b) Ausência: One saw that he was drunk (Viram que ele estava bêbado).
4. Tipo sintático da oração (oração principal ou encaixada);
- a) Oração principal: One shouldn't drink sour milk (Não se deve beber leite azedo).
 - b) Oração encaixada: What happens if one drinks sour milk? (O que acontece se beber leite azedo?)
5. Tipo de quantificação (universal ou existencial);
- a) Universal: In Spain they eat late (Na Espanha eles comem tarde).
 - b) Existencial: They're knocking at the door (Estão batendo na porta).
6. A individuação da ação (coletiva ou individual);
- a) Coletiva: They've raised the taxes (Eles elevaram os impostos).
 - b) Individual: They stole my bike yesterday (Roubaram minha bicicleta ontem).
7. A definitude (Definido ou indefinido);
- a) Definido: They've raised the taxes (Eles elevaram os impostos).
 - b) Indefinido: Here they've eaten seafood (Aqui comem frutos do mar).
8. A inclusão ou exclusão do falante;
- a) Inclui: In France you eat snails (Na França, você come caracóis).
 - b) Não inclui: In France they eat snails (Na França, eles comem caracóis).
9. A abrangência do conjunto de referentes (singular ou plural).

- a) Singular: They're knocking at the door. It's your mother (Estão batendo na porta. É a sua mãe).
- b) Plural: They've surrounded us (Nos cercaram).

Gast afirma que, no decorrer da análise, foi preciso mudar três perspectivas, já que os resultados mostraram algumas divergências em relação ao esperado:

- As formas verbais não finitas e as formas nominais são mais proeminentes como estratégia de impessoalização, superando, em línguas como o inglês, as estratégias com pronomes impessoais;
- A especificidade das variáveis, já que as estratégias variam bastante em relação aos parâmetros aos quais são sensíveis. Segundo o autor, o grau dessa diferenciação foi maior que o esperado;
- A questão de inclusão/exclusão do falante é, na verdade, uma questão de empatia, e não de referencialidade, já que há inclusão do falante quando há empatia entre o falante, o ouvinte e os demais membros de uma categoria.

De acordo com o estudo de Negrão e Viotti (2008), as construções impessoais são, no caso do PB, resultantes da forma como a variante brasileira do português conceitualiza os eventos e situações e a forma como envolve o falante/ouvinte nesses eventos. Essa forma de lidar com o sujeito está diretamente relacionada às mudanças sintáticas e semânticas ocasionadas pela participação das línguas africanas na formação do PB, uma vez que difere consideravelmente da variante europeia (NEGRÃO E VIOTTI, 2008, p.3):

Essa nossa ideia pode ser assemelhada à noção de *sémantaxe* proposta e desenvolvida por Manessy (1995), segundo a qual a estrutura sintática reflete meios de categorizar experiência e de organizar a informação. Esses meios são determinados culturalmente e adquiridos pelo contacto social e pela experiência. De acordo com o autor, a fala de diferentes povos que compartilham certos valores culturais pode adquirir um tipo de semelhança de família, que não pode ser explicada por nenhuma herança linguística, nem pelo uso dos mesmos processos gramaticais. Essa semelhança de família vem de uma visão de mundo comum, ou de meios comuns de categorizar a experiência.

Apesar de relativamente semelhantes, os processos de indeterminação e impessoalização guardam algumas diferenças, que estão, segundo os autores, na ausência/presença de conteúdo semântico no sujeito: enquanto as construções indeterminadoras apresentam sujeitos semanticamente significativos, ainda que desconhecidos e não realizados foneticamente, as construções impessoais não apresentam

traço de agentividade, sendo construídas com a categoria do sujeito vazia (09), (10), (11), ou com a presença de sujeitos expletivos com forma fonológica e sem conteúdo semântico (12), (13) (NEGRÃO; VIOTTI, 2008):

(9) Choveu muito esse mês.

(10) Tá fazendo muito calor em Salvador

(11) Hoje à noite vai ter uma festa na casa da Cecília.

(12) They say it will rain toninght (Disseram que vai chover essa noite).

(13) There is a lot of people in the square (Existe um grande número de pessoas na praça).

A forma de lidar com os vários tipos de sentenças tanto pessoais quanto impessoais é, segundo as autoras, a construção de um *continuum*, sendo que em um extremo estariam as sentenças com sujeito ativo, e no outro as sentenças impessoais. A indeterminação do sujeito estaria, então, localizada no meio do *continuum*.

A análise de Cabredo Hofherr (2003) segue parâmetros diferentes das mencionadas anteriormente. A autora, que analisa somente construções com verbo na terceira pessoa do plural, afirma que, ao contrário do que afirmam a maioria dos trabalhos que abordam a tipologia dos pronomes impessoais, os pronomes de terceira pessoa do plural sem antecedente não são um fenômeno unitário e não devem ser sempre tratados como arbitrários. Sendo assim, ela propõe cinco leituras para esses pronomes:

1. Leitura existencial específica: Se o evento está ancorado em um ponto específico do tempo.

Ex.: Tocan a la puerta.

2. Leitura existencial vaga: O evento está ancorado em um ponto menos específico do tempo.

Ex.: They have found a motorbike in the courtyard (Encontraram uma moto no pátio).

3. Leitura existencial inferida: Há inferência de resultado.

Ex.: Aquí han comido mariscos.

4. Leitura corporativa: Pressupõe um grupo particular como sujeito.

Ex.: Volvieron a aumentar el IVA.

5. Leitura universal: É possível com todos os tipos de predicado, mas depende de uma expressão locativa.

Ex.: En España hablan español.

A autora conclui que situações como as 1, 2 e 3 assemelham-se mais a indefinidos, enquanto as descritas em 4 e 5 são mais definidos, o que a leva a afirmar que existem dois tipos de referência anterior nos contextos impessoais. Assim sendo, pode se afirmar que essa é uma análise pautada em critérios puramente semânticos.

As análises mencionadas acima nos permitem afirmar que a diferença entre impessoalização e indeterminação é muito mais contextual que sintática, sendo que não existem estratégias exclusivas para uma ou outra condição, e sim contextos semânticos que favorecem mais uma que a outra. Assim, consideraremos, aqui, que tratam-se não de fenômenos tipologicamente diferentes, mas sim de um mesmo fenômeno que ocorre em níveis e contextos diferenciados. Por isso, nos respaldaremos na análise de níveis de indeterminação de Milanez (1982) (cf. seção anterior).

Iremos, a seguir, apresentar a descrição do fenômeno da indeterminação de acordo com algumas gramáticas normativas e, em seguida, apresentar algumas análises científicas recentes do tema.

2.2 A TRADIÇÃO GRAMATICAL

Por muito tempo – e, por muitos, ainda nos dias de hoje – acreditou-se que as gramáticas ofereciam uma completa descrição dos fenômenos linguísticos, trazendo em si concepções definitivas acerca da língua, as quais deveriam ser seguidas em prol da correção e bom funcionamento da comunicação. Sabemos, entretanto, que essa não é uma concepção adequada: a língua é um fenômeno vivo, em movimento, e a ciência que a descreve e analisa – a Linguística – é tão móvel quanto a língua, estando sempre em constante desenvolvimento.

É importante admitir, porém, que as gramáticas tradicionais prestaram, e ainda prestam importantes esclarecimentos acerca da descrição dos fenômenos linguísticos, por mais que essa descrição, na maioria dos casos, não esteja em total acordo com a realidade da língua. Por isso, julgamos importante comentar aqui a forma como o fenômeno da indeterminação é tratado nessas gramáticas.

No âmbito das gramáticas tradicionais, a análise do fenômeno permanece restrita às estratégias tradicionalmente conhecidas, sendo que pouquíssimos autores consideram as possibilidades inovadoras que compõem o fenômeno, ou o fazem em forma de notas de rodapé. Outra lacuna das gramáticas tradicionais diz respeito à parte semântica do fenômeno, que é consideravelmente importante para a análise.

Apesar de não estarmos propondo, aqui, uma análise no âmbito da gramática tradicional, consideramos relevante estabelecer algumas considerações acerca da abordagem da indeterminação do sujeito a partir deste ponto de vista, já que críticas ao modelo tradicional de análise podem servir de subsídio para uma abordagem científica. Para tanto, comentaremos, aqui, a definição de indeterminação do sujeito trazida por algumas gramáticas recentes e de grande circulação. São elas FERREIRA (2003), CUNHA & CINTRA (2008), e ROCHA LIMA (2013).

O que podemos perceber em comum nas três perspectivas é que, inicialmente, os autores fazem uma distinção bastante recorrente, traçando diferenças entre o sujeito determinado elíptico (oculto) e o sujeito indeterminado. Para os autores, o sujeito é oculto, mas determinável, quando a desinência verbal possibilita a retomada desse sujeito ou pelo contexto da enunciação, como nos exemplos abaixo:

(14) Como aceitaremos a existência de vida nos outros planetas?

(15) Ficamos um bocado sem falar.

(16) O velho vaqueiro rememorava as viagens / que fizera pelos sertões.

(17) Guilhermina bocejou. Iria adormecer? Pôs-se a calcular as horas.

Segundo os autores, em (14) e (15) o sujeito elíptico *nós* é identificável pela desinência *-mos*. Há, também, importantes diferenças semânticas entre os dois exemplos, que são ignoradas pela classificação feita pelo autor: em (14), percebe-se que há uma conjectura que não apresenta o mesmo valor de verdade da afirmação feita em (15). Há, também, diferenças no que diz respeito à especificidade dos sujeitos e à inclusão/exclusão do falante no contexto de fala. Em (16) e (17) o sujeito oculto é identificável pelo contexto (O velho vaqueiro e Guilhermina, respectivamente).

Em relação ao sujeito indeterminado, temos definições também semelhantes. De acordo com Ferreira (2003), “Uma oração tem sujeito indeterminado quando o falante que a constrói

não quer – ou não pode – fixar o sujeito”. Assim sendo, a indeterminação do sujeito é possível através das seguintes estruturas sintáticas:

1. Verbo na terceira pessoa do plural.

(18) *Dizem* que pintar é uma boa distração.

2. Verbo na terceira pessoa do singular + pronome *se*.

(19) *Era-se* mais feliz naquele tempo.

Cunha & Cintra (2008), sobre a indeterminação do sujeito, afirmam: “Dizemos que o sujeito é indeterminado quando o verbo não se refere a uma pessoa determinada, ou por se desconhecer quem executa a ação, ou por não haver interesse no seu conhecimento”. As possibilidades dadas pelos autores para se indeterminar o sujeito são as mesmas anteriores: verbo na terceira pessoa do plural ou na terceira pessoa do singular + o pronome *se*.

Rocha Lima (2013) segue a mesma linha dos autores citados anteriormente, afirmando que os sujeitos são indeterminados quando não queremos ou não podemos determinar o agente. As estratégias mencionadas para a indeterminação são as mesmas, com uma ressalva relativa à transitividade verbal:

- usa-se o verbo na terceira pessoa do singular acompanhado da partícula *se*, desde que o verbo seja intransitivo, ou traga complemento preposicional. (ROCHA LIMA, p.289)

O que podemos perceber é que há, entre as duas definições (sujeito oculto e sujeito indeterminado), uma diferença de classificação que, em termos práticos, precisa ser revista. A diferença entre essas duas formas de sujeito está nos graus de definição desses sujeitos, não sendo conceitual. Sobre isso, Dias e Ladeira (2013, p.94) afirmam que

O preenchimento ou não do lugar do sujeito indeterminado estaria em função do grau de indeterminação que se revela como efeito do enunciado com um todo. Daí a nossa hipótese de que existem enunciações em que pronomes e expressões deixam as sentenças + indeterminadas e outras - indeterminadas. Dessa maneira, podemos delinear categorias de indeterminação, de acordo com os seus modos de enunciação.

Temos, portanto, casos diferentes de virtualização e projeção da identidade dos sujeitos, sem que haja necessariamente uma diferença nas classificações. A real diferença estaria, portanto, no nível em que essa projeção acontece.

Existem outros problemas relacionados à concepção da gramática tradicional. O primeiro deles diz respeito à pequena quantidade de possibilidades de se indeterminar o

sujeito, que se restringe, na gramática normativa, a apenas duas formas, enquanto a realidade da língua nos mostra que o fenômeno é muito mais amplo, incluindo várias outras possibilidades.

2.3 ABORDAGENS VARIACIONISTAS

2.3.1 No português brasileiro

Em seu trabalho acerca do funcionamento dos sujeitos indeterminados no português falado nas três capitais do sul, Setti (1998) busca verificar quais são as formas de indeterminação mais recorrentes no português oral. Para tanto, a autora considerou os fatores sociais Localidade (Curitiba, Florianópolis e Porto Alegre), Faixa etária (de 25 a 50 anos e mais de 50 anos), Sexo (feminino e masculino) e escolaridade (primária, ginásial e secundária), encontrando como resultado as seguintes formas:

- a. a gente
- b. eles
- c. eu
- d. nós
- e. se
- f. tu
- g. você(s)
- h. voz passiva sem agente
- i. voz passiva sintética
- j. Ø + verbo na terceira pessoa do singular
- k. Ø + verbo da terceira pessoa do plural
- l. algumas formas nominais, como o camarada, o indivíduo, a pessoa, as pessoas, o pessoal, o sujeito e o cara.

Dentre as constatações às quais a autora chegou, vale destacar a diferenciação da capital Curitiba em relação às outras duas capitais do sul. Segundo os resultados obtidos pela autora, Curitiba é a capital que mais indetermina o sujeito, e que dentre as três é a única que não utiliza o pronome *tu* como indeterminador, dando preferência ao pronome *você*. Em relação ao uso das variantes em cruzamento com a variável Faixa etária, os resultados mostraram que os falantes com menos de 50 anos tendem a utilizar com maior frequência o pronome *você* em detrimento do pronome *tu*, o que mostra que *você* é uma estratégia inovadora. Um dado

importante em relação aos resultados obtidos pela autora diz respeito à diminuição do uso da passiva sintética como estratégia indeterminadora, o que já estava sendo verificado em outras variedades linguísticas.

Souza (2007), estudando a fala da cidade de Belo Horizonte, analisa o percurso do pronome *eles* como estratégia de indeterminação do sujeito, enfocando esse pronome tanto na sua forma plena quanto na sua forma nula, representada pela flexão verbal na terceira pessoa do plural. A amostra utilizada por ela é composta por 36 informantes (18 homens e 18 mulheres), estratificados em três faixas etárias (faixa I – 15 a 29 anos, faixa II – 30 a 45 anos e faixa III – 46 a 76 anos), divididos em dois níveis de escolaridade: baixa escolaridade (ensino fundamental) e alta escolaridade (ensino médio e superior). As variáveis linguísticas consideradas pela autora foram:

1. Referência (+/- determinada) da variável dependente
2. Referência do SN que é retomado pelas variantes pronominais
3. Concordância de gênero entre o SN e a variável dependente
4. Concordância de número entre o SN antecedente e a variável dependente
5. Traço [\pm humano] do sujeito
6. Tipo de oração

As frequências obtidas para cada uma dessas estratégias foram as seguintes:

Tabela 4 - Distribuição das ocorrências de terceira pessoa na fala de Belo Horizonte

Estratégia	Número de ocorrências	Percentual	Total
Eles (pleno)	17	74%	239
Ø + terceira pessoa do plural	62	25%	239

A autora fez importantes constatações sobre o uso do *eles* como recurso de indeterminação do sujeito, principalmente no que diz respeito aos níveis de indeterminação estudados por ela. Ela concluiu que este item está se tornando cada vez mais abstrato, já que está perdendo conteúdo referencial. A realização plena do pronome é um dado significativamente relevante em relação à análise no *corpus* de Salvador, já que os resultados se mostraram opostos: enquanto, em Belo Horizonte, a realização plena é favorecida, em Salvador ela é desfavorecida, conforme veremos posteriormente.

Em um estudo comparativo entre a língua escrita veiculada em jornais do Rio de Janeiro e as falas culta e popular que compõem o *corpus* dos projetos NURC-RJ e PEUL, Duarte (2007) analisa tanto sujeitos determinados de terceira pessoa, os chamados sujeitos de referência definida, quanto os sujeitos de referência arbitrária, os sujeitos indeterminados, em sentenças finitas, comparando a fala de informantes com curso superior completo com a dos que cursaram apenas o ensino fundamental e o médio. Em seguida, a autora confronta os resultados obtidos aos do português europeu falado e escrito, recolhidos da amostra de Duarte (1995). Para os sujeitos de referência definida, a autora apresenta os seguintes resultados:

Tabela 5 - Corpora orais – sujeitos expressos de terceira pessoa

Estratégia	PE			PB (NURC)			PB (PEUL – 1980)			PB (PEUL – 2000)		
	Freq.	Total	%	Freq.	Total	%	Freq.	Total	%	Freq.	Total	%
3ª p. s.	80	285	28	254	419	61	1244	1545	81	1082	288	81
3ª p. p.	34	132	26	83	127	65	274	339	81	303	369	82

Tabela 6 - Corpora escritos – sujeitos expressos de terceira pessoa

PE	PB
17/244 (7%)	122/241

Para os corpora de sujeito de referência arbitrária, os resultados são os seguintes:

Tabela 7 - Corpora orais – sujeitos indeterminados

Variedade	Se		Eles		A gente		Nós		Você		Tu		Zero		Total
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	
PE	36	38	20	22	18	20	12	14	5	6	--	--	--	--	91
PB (NURC)	26	8	50	16	41	13	8	2	140	44	--	--	56	17	321
PB (PEUL - anos 80)	18	2	104	13	117	15	11	1	391	49	6	0,7	152	19	799
PB (PEUL - anos 2000)	11	2	84	3	131	21	31	5	284	45	14	2	74	12	629

Tabela 8 - *Corpora* escritos – sujeitos indeterminados

Variedade	Você		Eles		A gente		Se		Nós		Total
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	
PE	--	--	11	4	--	--	181	69	72	27	264
PB	7	3	29	11	13	5	97	36	122	45	268

Essa análise forneceu importantes dados no que diz respeito à forma como a gramática da fala se implementa na escrita e das diferenças de conhecimento linguístico entre indivíduos letrados e não-letrados. A autora pôde constatar, também, que ainda é relevante a distância entre o português brasileiro falado e escrito e entre as variantes portuguesa e brasileira.

Em 2008, Duarte realizou uma análise semelhante à anterior, mas em sentenças infinitivas, como as descritas nos exemplos abaixo. O *corpus* da análise foi composto por 12 entrevistas do projeto NURC–RJ e por entrevistas extraídas do *corpus* de Cresti e Moneglia (2005) para o português europeu. Para o *corpus* escrito, foram utilizados artigos de opinião e crônicas publicados em jornais de Lisboa e em jornais cariocas. Abaixo, alguns exemplos recolhidos do *corpus* (DUARTE, 2008, p.11):

(20) É muito difícil [Ø conseguir uma carga de quarenta horas], você sabe disso, te obriga a ir a vários colégios, deslocamento, passagem.

(21) É uma das cidades que tem o melhor, melhor é, padrão de vida, ou seja, uma das melhores cidades do mundo [pra *se* morar].

(22) Outro dia eu tive, tive que, expulsar um aluno de sala, realmente é, triste né, [*você* ter que expulsar um aluno de sala].

(23) É muito gostoso [*a gente* analisar isso], é muito bonito.

Os resultados obtidos pela autora foram os seguintes:

Tabela 9 - Distribuição dos dados – sentenças infinitivas

Variantes	Modalidade							
	Fala				Escrita			
	PE		PB		PE		PB	
	F	%	F	%	F	%	F	%
Nulo	68	91	156	78	133	94	133	90
Se	6	8	3	1,5	9	6	10	7
A gente	1	1	4	2	--	--	1	1,5
Você	--	--	37	18,6	--	--	1	1,5
Total	75		200		142		147	

Na análise dos resultados, a autora conclui que há considerável proximidade entre o português europeu falado e escrito no que diz respeito às sentenças infinitivas e o caráter conservador da escrita brasileira. Os dados acima também mostram que na fala e na escrita formais o português brasileiro e o europeu apresentam gramáticas muito próximas.

O estudo de Duarte trouxe novas e relevantes constatações acerca do alçamento de pronomes referenciais para a posição de sujeito em construções impessoais. Para a autora, línguas como o português brasileiro, com orientação para o discurso, não costumam apresentar elementos sem conteúdo semântico, como os expletivos.

A análise feita por Campos (2010) busca analisar o funcionamento dos mecanismos de indeterminação do sujeito que eram recorrentes no final do século XIX, utilizando como *corpus* 38 cartas escritas nesse período. Os resultados obtidos pelo autor foram os seguintes:

Tabela 10 - Distribuição das ocorrências de indeterminação do sujeito em cartas do século XIX

Estratégia	Frequência	Percentual	Total
Se	169	48%	355
1ª pessoa do plural	116	33%	355
Voz passiva analítica	40	11%	355
3ª pessoa do plural	30	8%	355

Estudos dessa natureza são muito importantes para que se tenha uma dimensão do funcionamento do fenômeno em sincronias pretéritas da língua, sendo que os resultados podem lançar luz sobre a forma como determinadas variáveis passaram a ser utilizadas. A dificuldade em se estudar sincronias passadas, entretanto, está na dificuldade de se obter dados de fala de épocas remotas, já que não se dispõe de gravações de fala dessas épocas. É, portanto, necessário desenvolver estratégias metodológicas que possibilitem essa análise, e muitos autores optam por estudar cartas e peças de teatro da época por serem essas as formas escritas mais próximas da fala.

O próximo exemplo de análise variacionista é a de Coan (2012), sobre as construções com o pronome indeterminador *se* em anúncios publicitários de revistas paulistanas, no início do século XX. É importante ressaltar que, em sua análise, o autor considera não somente aos elementos verbais, mas também ao contexto sócio-histórico e cultural em que o *corpus* foi produzido. A seguir, algumas ocorrências destacadas pelo autor:

(24) Aceitam-se encomendas. Executam-se obras em marmores, ladrilhos, azulejos, de qualquer importância.

(25) Aceita qualquer classes de trabalhos a oleo, pastel, aquarella, crayon, etc. Especialidade em ampliações. Attendem-se a chamados do interior.

(26) Aceitam-se pedidos á Rua Tupinambá n. 10 e Galeria de Christal n. 15.

(27) Compra-se ou troca-se por outros artigos: Dentes avulsos, Retalhos de platina e Dentaduras. Aprompta-se com perfeição Receitas de Oculistas.

(28) Vendas por atacado e a varejo. Recebe-se por todos os paquetes as últimas novidades.

O autor interpreta essas ocorrências como sempre correlacionadas a um tópico previamente expresso, sendo que ocorre a coindexação a um sujeito nulo. Assim sendo, nos contextos dessas propagandas, o *se* aponta para o traço humano e indeterminado do agente.

Esse dado, assim como os dados obtidos pelos estudos de Duarte, aponta para uma mudança em relação as passivas sintéticas, que passam a assumir valor indeterminador.

A última análise que servirá de exemplo de abordagem do tema no âmbito da linguística moderna foi realizada por Lucchesi (2015), tendo como *corpus* amostras de fala recolhidas em quatro comunidades situadas no interior da Bahia, *corpus* que o autor chama de português afrobrasileiro. Para realizar essa análise, foram consideradas as seguintes estratégias de indeterminação do sujeito:

- (i) partícula *se*
- (ii) nós
- (iii) a gente
- (iv) você
- (v) eles ou *-mos*
- (vi) forma verbal não marcada da 3ª pessoa do singular

O autor afirma que estudar comunidades afrobrasileiras isoladas, originadas de antigos quilombos e compostas majoritariamente por descendentes de escravos, pode fornecer importantes indícios acerca da forma como a língua portuguesa foi nativizada nessas comunidades, já que as mudanças pelas quais a língua passou foram conservadas por muito tempo devido ao nível de isolamento dessas comunidades, o que fez com que essa variedade linguística apresentasse características que a diferenciam até mesmo de outras variedades rurais do português brasileiro. Essa situação que só se alterou na primeira metade do século XX, com a influência do padrão culto sobre as variedades rurais do português brasileiro. (LUCCHESI et al, 2009, p.76).

Por se tratar de uma análise também pautada em nas características de formação do PB, este estudo fornece dados extremamente relevantes ao que é proposto aqui, visto que contribui para a compreensão da relevância do contato entre línguas para o fenômeno da indeterminação.

2.3.2 No português moçambicano

Veremos, agora, alguns estudos do funcionamento dos sujeitos tanto definidos quanto indeterminados em situações sociolinguísticas semelhantes às verificadas no Brasil, durante a formação do português brasileiro. Tem-se, fora do Brasil, situações linguísticas relativas ao

uso do português que se assemelham consideravelmente ao que se passou aqui: localidades para onde a língua portuguesa foi transplantada em ocasião da expansão marítima de Portugal. Algumas dessas regiões desenvolveram variedades distintas de língua – as chamadas línguas crioulas – como é o caso de Cabo Verde, Guiné-Bissau e São Tomé e Príncipe, enquanto outras regiões desenvolveram variedades diferentes da língua portuguesa, assim como ocorreu no Brasil: Angola e Moçambique. (PETTER, 2009)

Moçambique, país localizado no sudoeste africano, começou a ser colonizado pelos portugueses no final do século XV, mas o território só recebeu maior atenção da metrópole no final do século XIX, quando Portugal perdeu a influência sobre a nação brasileira. Sobre a fixação da língua portuguesa em território moçambicano, Oliveira (2016, p.14) afirma que

é importante ressaltar que o maior povoamento e uma preocupação de fixação europeia no território se deram somente no século XIX, após a perda do poder colonial sobre o Brasil. Logo, a língua portuguesa à qual os moçambicanos foram (e ainda estão) expostos é representativa do português contemporâneo falado na metrópole, de forma que o “ideal linguístico” parece ser bastante espelhado no falar europeu atual, diferentemente do que se vê no Brasil, em que levas e mais levas de portugueses foram se fixando ao longo de toda a história da colonização e do período imperial, trazendo as variadas fases do português falado na metrópole.

Atualmente, apesar de ter o português como língua oficial, apenas 8,8% da população utiliza plenamente a língua em ambiente doméstico, sendo a língua mais falada a *emakhua* e as línguas do grupo *xichamghana*, segundo Norte e Rios Neto (2008, apud OLIVEIRA, 2016, p.17).

Em sua dissertação de mestrado, Oliveira (2016) faz, sob o aporte teórico e metodológico do modelo de princípios e parâmetros (CHOMSKY, 1981) e da sociolinguística variacionista (LABOV, 1972), uma análise do funcionamento dos sujeitos no português moçambicano, incluindo tanto os sujeitos nulos quanto os sujeitos indeterminados, através da análise de entrevistas recolhidas em Maputo, capital de Moçambique. Para a análise, o autor considerou as seguintes variáveis linguísticas e sociais:

- (i) realização fonológica do sujeito
- (ii) categoria do sujeito
- (iii) presença de elemento preposto ao verbo
- (iv) tempo verbal
- (v) regularidade morfológica do verbo
- (vi) forma simples ou perifrástica do verbo
- (vii) marca indicativa de número e pessoa

- (viii) concordância verbal
- (ix) animacidade do sujeito
- (x) (in)determinação referencial do sujeito
- (xi) estatuto da oração quanto à sua (in)dependência sintática
- (xii) gênero
- (xiii) faixa etária
- (xiv) escolaridade do informante
- (xv) forma como o português foi adquirido

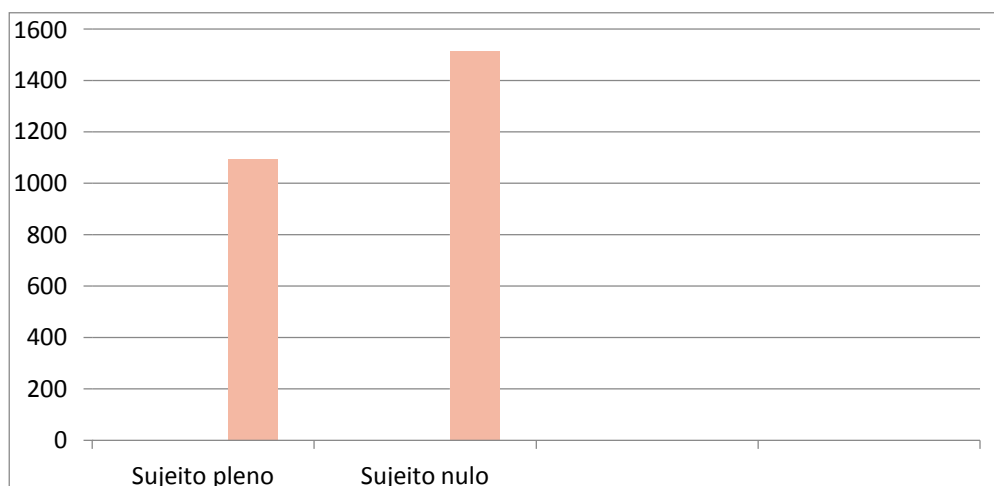
Os resultados encontrados para o sujeito nulo foram os seguintes:

Tabela 11 - O sujeito nulo em Moçambique

Pessoa verbal	Frequência de sujeitos nulos	Percentual	P.R.	Total
1ª pessoa do singular	577	54,8%	.48	1053
2ª q pessoa do singular	21	91,3%	.90	23
3ª pessoa do singular	425	53,9%	.41	789
1ª pessoa do plural	276	70,4%	.64	392
2ª pessoa do plural	215	49,5%	.56	336
Total	1514	--	--	2593

Se compararmos o total de sujeitos nulos e o total de sujeitos plenos, teremos o seguinte resultado:

Gráfico 2 - Sujeitos nulos e plenos



Fonte: OLIVEIRA, 2016

O que podemos concluir com esses resultados é que, como se esperava, o português moçambicano ainda apresenta norma semelhante à do português europeu, o que pode ser comprovado pelo alto índice de sujeitos nulos, ao contrário do que ocorre no português brasileiro, fato que também é atestado, de acordo com a discussão levantada pelo autor, pela alta taxa de concordância verbal observada.

Em relação ao sujeito indeterminado, o autor coletou ocorrências como as seguintes (p.67):

(29) Às vezes há dias que não Ø consegue nada.

(30) Aqui agora já não Ø faz mais isso.

(31) Não vejo diferença (...) vive Ø junto, é marido é mulher.

(32) Ø Tem que comprar muita coisa, ou então Ø compra um cabrito e diversas coisas que eles comem.

(33) Só quando *você* tem aquela classe pequena, mas nessa classe pequena você aproveitava.

(34) Quando *você* vai para a escola, *você* não vai aprender nada.

Como podemos ver nos exemplos acima, várias estratégias encontradas não são comuns na norma oficial de Moçambique, que é o português europeu, mas sim no português brasileiro. Para explicar esse fato, o autor afirma que é possível que a morfologia verbal rica do português moçambicano seja apenas superficial, pois, no português brasileiro, estratégias como as descritas acima surgiram justamente devido ao enfraquecimento da morfologia verbal. Sendo assim, a riqueza da morfologia verbal do português moçambicano pode ser associada apenas ao uso escolar, perdendo sua força em ambientes menos formais.

2.3.3 No português angolano

A situação de Angola, no que diz respeito à colonização portuguesa, seguiu caminhos muito semelhantes aos de Moçambique: a colônia só recebeu verdadeira atenção da metrópole depois da independência da colônia mais promissora - o Brasil. Dados recentes acerca do

português em Luanda afirmam que a maior parte da população é bilíngue (português e kimbundo), enquanto uma parte considerável é monolíngue em português e uma pequena parcela é falante apenas de quimbundo, segundo Mingas (2000, apud ARAUJO, 2010, p.6).

Com vistas a identificar as formas pronominais utilizadas para indeterminar o sujeito no português de Luanda, Pitombo e Lopes (2011) comparam o fenômeno com o que foi observado no português brasileiro. Em relação às frequências gerais dos recursos de indeterminação, as autoras apresentam os seguintes resultados, levando em consideração duas variáveis: falantes que têm o português como língua nativa e falantes que tem alguma língua nacional como língua materna.

Tabela 12 - Distribuição das ocorrências – português e língua nacional

Variante	Língua native			
	Português		Língua nacional	
	Frequência	Percentual	Frequência	Percentual
3ª pessoa do plural	31	14	31	20
Nós	109	49	36	23
Você	16	7	39	25
Se	37	16	29	18
Tu	28	12	7	4
A gente	--	--	13	8

Abaixo, exemplos de algumas ocorrências encontradas pelas autoras:

(35) Anteriormente *nós* criamos e *metemos* na escola.

(36) Eu penso que neste momento, nota-se um crescimento muito grande da mulher a nível educacional, nas universidades e em todos os níveis.

(37) Sim, já *dispararam* contra mim só que prontos, o fator sorte permaneceu.

(38) Quando *você* não está presente, o grupo recolhe os fascículos, passa a informação e *você* não pode reclamar.

Após comparar os resultados obtidos com o estudo feito por Cavalcante (1999) sobre o PB do século XIX, as autoras percebem que há uma semelhança entre as duas sincronias, principalmente no que diz respeito à preferência pelo uso da estratégia *nós* e pelo uso do *você* e do *a gente*, que são recorrentes no português brasileiro, mas pouco utilizados no português europeu. Com isso, as autoras percebem que o português angolano atual está passando por mudanças semelhantes às que foram observadas no português brasileiro do século XIX.

Estudos como este são fundamentais para que se possa compreender como o contato entre línguas afetou outras variedades do português, além da brasileira, o que pode corroborar os resultados obtidos no PB.

Nesta seção, comentamos a importância de se considerar a indeterminação como um fenômeno gradual, que se dá em níveis diferentes, e comentamos as diferenças epistemológicas entre conceitos que, por vezes, são tratados como sinônimos, mas que aqui consideramos como fazendo parte de um *continuum*: a indeterminação e a impessoalização do sujeito. Comentamos, também, algumas análises variacionistas do tema que foram realizadas no português brasileiro, angolano e moçambicano e que motivaram as nossas escolhas em relação às variantes que compõem o fenômeno e às variáveis linguísticas e sociais, avaliando a contribuição desses estudos para a análise realizada aqui.

2.4 O PARÂMETRO DO SUJEITO NULO

Os estudos realizados acerca do sujeito nulo têm sido realizados majoritariamente sob a luz da teoria de princípios e parâmetros, modelo proposto por Chomsky (1981) e que postula que todas as línguas naturais são dotadas de princípios, leis gerais aplicáveis a todas as línguas, e parâmetros, que são propriedades mais específicas que podem ou não ser exibidas por uma língua, sendo que a fixação dos princípios e parâmetros é o que compõe a gramática de uma língua. Dentre os parâmetros que fazem parte de uma língua e que são passíveis de estudo, trataremos, aqui, do chamado parâmetro do sujeito nulo.

Os pressupostos gerativistas relativos ao Parâmetro do Sujeito Nulo defendem que algumas línguas permitem que uma categoria nula (a chamada categoria *pro-drop*) ocupe a posição de sujeito nas sentenças finitas, desde que a morfologia verbal seja rica o suficiente para recuperar a referência desse sujeito. Temos, portanto, que o licenciamento para o sujeito nulo está intimamente relacionado com a vivacidade morfológica da língua. Kaiser (2006,

p.14), sobre essa relação, retoma alguns postulados da teoria gerativa, afirmando que:

De acordo com Chomsky (1981, 1982), que propôs pela primeira vez o estabelecimento do parâmetro *pro-drop*, a diferença entre línguas *pro-drop* e línguas não-*pro-drop* tem a ver com a capacidade ou incapacidade da categoria funcional da concordância de identificar ou tornar recuperável, por meio de traços morfológicos, o sujeito nulo. Além disso, acredita-se que um sujeito nulo precise ser formalmente licenciado. Ou seja, a ocorrência de um sujeito nulo, sendo representado pela categoria vazia *pro*, e determinada não só por um processo de identificação, mas também por um processo de licenciamento.

Sendo assim, línguas que permitem o sujeito nulo são as línguas com marcação positiva para o parâmetro *pro-drop* (+*pro-drop*), desde que estas, além de apresentar a categoria vazia, identifiquem e licenciem o conteúdo nominativo através de traços de número e pessoa do verbo (39) (40), e línguas que não permitem a categoria vazia como sujeito seriam as línguas com marcação negativa para o parâmetro *pro-drop* (-*pro-drop*) (41):

(39) Ø Fui ao mercado ontem.

(40) Ø Parlo molto poco inglese.

(41) I bought lots of fruits.

*Bought lots of fruits.

A correlação entre a morfologia verbal e o licenciamento ou não licenciamento do sujeito nulo é verificável em várias línguas que reconhecem o parâmetro *pro-drop*, como é o caso do francês moderno, por exemplo. A língua passou por sucessivas mudanças na morfologia flexional (Tabela 13), que, com o passar do tempo, passou de seis formas distintas para apenas duas ou três. Essa mudança fez com que os sujeitos nulos se tornassem menos frequentes na língua.

Tabela 13 - Enfraquecimento da morfologia verbal no francês

Presente do indicativo	Francês antigo		Francês médio		Francês modern		
1ª p.s.	chant	∅	chant	[ò]	jê	chant	∅
2ª p.s.	chant	es	chant	[ò]	tóóóu	chant	∅
3 p.s.	chant	e(t)	chant	[ò]	il/elle	chant	∅
1ª p.p.	chant	ons	chant	[õ]	nous	chant	[õ]
					on	chant	∅
2ª p.p	chant	ez	chant	[e]	vous	chant	[e]
3ª p.p	chant	ent	chant	[ò]	ils/elles	chant	∅

Fonte: KAISER (2006, p.23)

Amplios estudos acerca do tema afirmam que o português brasileiro, assim como o francês, estaria deixando de ser uma língua “evite pronome” (como também é o caso do italiano, língua de sujeito nulo prototípica no grupo românico), já que a falta de especificação da pessoa do discurso através da morfologia flexional induz o falante a realizar o sujeito com mais frequência, fenômeno que, por sua vez, seria motivado pela inserção dos pronomes *a gente* e *você* no quadro pronominal do português brasileiro:

A perda da propriedade de licenciar o sujeito referencial nulo relaciona-se diretamente com o enfraquecimento do paradigma da flexão de número e pessoa do verbo, pois Duarte (1993) demonstra, em um estudo diacrônico também focalizando a norma culta, que o nível de preenchimento do sujeito pronominal, no seu *corpus* de textos escritos para peças teatrais, salta de 25%, em 1918, para 46%, em 1937; atingindo os níveis de 67% e 74% de sujeitos realizados, em 1975 e 1992, respectivamente. Para a autora, a elevação do nível de realização do sujeito pronominal 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 3ª pessoa do singular, ou seja, sem um morfema específico de pessoa e número. Com isso, em alguns tempos verbais, como no imperfeito do indicativo, a flexão verbal restringe-se a apenas duas pessoas, como se pode ver no Quadro 1. Tal enfraquecimento da flexão verbal estaria, portanto, na base do enfraquecimento da propriedade de licenciar o sujeito referencial nulo no PB. (LUCCHESI et al, 2009, p. 173)

Esse processo fica evidenciado no português brasileiro quando comparamos o nosso sistema pronominal com o do português europeu, como pode-se observar na tabela 14:

Tabela 14 - Enfraquecimento da morfologia verbal no português

Presente do indicativo	Português europeu				Português brasileiro			
	1ª p.s.	Eu	cant	--	o	Eu	cant	--
2ª p.s.	Tu	cant	a	s	Você	cant	a	Ø
					tu	cant	a	Ø
3ª p.s.	ele/ela	cant	a	Ø	ele/ela	cant	a	Ø
1ª p.p.	Nós	cant	a	mos	nós	cant	a	mos/ Ø
					a gente	cant	a	Ø
2ª p.p.	Vós	cant	a	is	vocês	cant	a	m/ Ø
	vocês	cant	a	m				
3ª p.p.	eles/elas	cant	a	m	eles/elas	cant	a	m/ Ø

Fonte: KAISER (2006, p.23)

Observa-se, nos dados da tabela acima, que o português europeu apresenta um paradigma verbal composto por seis formas, enquanto o português brasileiro culto apresenta apenas quatro, e o português popular, apenas duas. Como dissemos anteriormente, esse enfraquecimento da morfologia verbal pode ser diretamente associado à inserção dos pronomes *a gente* e *você* no quadro pronominal do português brasileiro, evento que teve como consequência a diminuição da frequência dos sujeitos nulos.

A influência da diminuição da frequência dos sujeitos nulos também é verificável nos sujeitos indeterminados. Duarte (1995) buscou verificar como funciona essa influência, acreditando que a frequência dos sujeitos indeterminados pronominais realizados foneticamente também passaria a ser maior, em um claro processo de encaixamento da mudança. Em números absolutos, a autora constatou, no *corpus*, a presença de 65% de sujeitos indeterminados (ou sujeitos de referência arbitrária) realizados foneticamente, o que, *a priori*, confirma a sua hipótese inicial. Na análise aqui realizada, também verificaremos se os sujeitos indeterminados foneticamente realizados são mais frequentes, considerando que o português brasileiro está comprovadamente perdendo a obrigatoriedade de se evitar o sujeito foneticamente expresso.

Nesta seção, comentamos a relação entre o parâmetro do sujeito nulo e a indeterminação do sujeito, considerando que há forte relação entre os dois fenômenos, uma vez que mudanças na realização do sujeito pronominal podem ter afetado o funcionamento dos mecanismos de

determinação do sujeito. Pretendemos, através das análises da variável Realização fonética do sujeito indeterminado, verificar a validade dessa hipótese.

3 TEORIA E MÉTODO

Descreveremos, aqui, quais os conceitos e as diretrizes teóricas e metodológicas que guiarão o estudo realizado, tanto no que diz respeito ao estabelecimento de um conceito de língua adequado ao nosso objetivo, quanto em relação ao estabelecimento da variável dependente e das variáveis independentes que serão estudadas. A metodologia que será apresentada aqui é tradicional nos estudos sociolinguísticos, sendo que há um grande número de manuais sobre o assunto publicados por linguistas pioneiros no uso desse método.

3.1 A ANÁLISE VARIACIONISTA

A sociolinguística se constitui em uma importante vertente de análise das línguas, tanto no que diz respeito à explicação dos fatos linguísticos em uma perspectiva científica quanto em relação ao surgimento de políticas de ensino de línguas que busquem incluir em suas pautas a heterogeneidade inerente às mesmas. Nesse sentido, a teoria sociolinguística busca entender e explicar a língua tanto como fato social, já que ela não funciona de modo dissociado da realidade das sociedades às quais serve, quanto como submetida à variabilidade e pluralidade, sendo, ao mesmo tempo, produto e meio de expressão de cultura. Nas palavras de Meillet (1906, p. 17, apud LABOV et al, 2006)

A língua é uma instituição com autonomia própria; devem-se determinar, portanto, as condições gerais de desenvolvimento a partir de um ponto de vista puramente linguístico; [...] mas como a língua é [também] uma instituição social, disso decorre que a linguística é uma ciência 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 – às vezes imediatas e, no mais das vezes, mediatas e indiretas.

Toda comunidade de fala é composta por diversos fenômenos em variação, seja no âmbito fonológico, morfológico, morfossintático, lexical, semântico, etc. A esses fenômenos chamamos de variáveis linguísticas, cujo delineamento é condicionado por fatores internos e externos ao sistema linguístico – as variáveis independentes. As variantes são, então, o feixe de possibilidades formais que, juntas, formam o fenômeno que se encontra em variação, ou ainda, segundo Labov (1975, p.188), “Variantes linguísticas são, portanto, diversas formas de se dizer a mesma coisa em um mesmo contexto, e com o mesmo valor de verdade”. Labov (1972) menciona três critérios indispensáveis para que se estude adequadamente determinado fenômeno como variável linguística:

Definir o número exato de variantes que formam o fenômeno em questão;

Estabelecer toda a multiplicidade de contextos em que ela aparece;

Elaborar um índice quantitativo que permita medir os valores das variáveis.

As variantes linguísticas podem coexistir harmonicamente, em um processo denominado variação estável (quando não há um movimento de mudança a favor de uma ou outra variante, como no caso da variação *nós x a gente* no português brasileiro) e podem estar em disputa, sendo que uma das variantes se sobrepõe às outras, como no caso da transição *vossa mercê > você*, no português (mudança em curso).

Em última análise, afirmamos que a sociolinguística fundamenta-se na premissa de que a natureza do fenômeno linguístico não permite um enfoque somente estrutural, sendo a análise social igualmente relevante. A língua é dotada de heterogeneidade estruturada e, por isso, regular, sistematizável.

Surgindo em confronto à teoria estruturalista, que, por não considerar o caráter social da língua, não foi capaz de explicar os motivos pelos quais as línguas mudam, a sociolinguística quantitativa/sociolinguística variacionista (LABOV, 1972; WEINREICH, LABOV e HERZOG, 1968). Esse modelo teórico-metodológico de análise linguística empírica que tem como mentor o linguista norte-americano William Labov, tendo como objetivo estudar quais são os padrões linguísticos de determinada comunidade de fala, observando quais são as formas linguísticas que se encontram em variação naquela comunidade e sistematizando quantitativamente os fatores que motivam essa variação. A variação é o pré-requisito básico para o funcionamento pleno da língua, e é motivada pela heterogeneidade do repertório de vivências e características sociais dos falantes que a utilizam. Assim sendo, os estudos da sociolinguística variacionista buscam comprovar que a variação das línguas não é aleatória e, por isso, passível de sistematização. Aplicar o método da sociolinguística variacionista é, em linhas gerais, escrutinar a língua em busca das suas regras variáveis, das leis que regem a variação.

Também faz parte do escopo da sociolinguística a análise da estabilidade de cada variante, com vistas a analisar se determinado fenômeno se encontra em variação ou em um processo de mudança linguística em curso, considerando sempre as características extralinguísticas da comunidade analisada, como nível socioeconômico, sexo, idade, entre outros. De acordo com a sociolinguística variacionista, a variabilidade das línguas não se constitui em um empecilho para o seu funcionamento, mas sim o contrário: um sistema homogêneo não seria funcional (BORTONI-RICARDO, 2014, p.52). Sendo assim, é

possível escrutinar, estatisticamente, a influência dos fatores internos e externos à estrutura da língua para o fenômeno em variação e para uma eventual mudança linguística.

A motivação social dos fenômenos de variação e mudança fica ainda mais evidente se considerarmos os conceitos de prestígio, estigma, inovação e conservadorismo, combinando tais fatores com as variáveis sociais dos informantes (MONTEIRO, 2000, p.66). Tanto a noção de prestígio quanto a de estigma linguístico não estão relacionados à estrutura da língua, mas sim aos falantes que utilizam a forma linguística em questão. Enquanto aquelas formas utilizadas por pessoas de maior status social, seja por questões econômicas ou de qualquer outra ordem, são mais valorizadas pela comunidade de fala, as variantes que sofrem estigma não gozam do mesmo status social. Monteiro (2000, p.65), sobre as variantes estigmatizadas, afirma que

Uma vez que a variação linguística pressupõe a valoração social, as variantes empregadas por falantes dos estratos mais baixos da população em grande parte são estigmatizadas. E o preconceito é tanto mais forte quanto maior for a identificação da forma com a classe discriminada. À proporção que passa a ser usada por outros grupos, o estigma vai diminuindo até deixar de existir completamente, se a variante é aceita pela classe dominante. É o que está acontecendo com o emprego do *ele* em função de objeto direto.

Em relação aos conceitos de variantes conservadoras e inovadoras, temos uma situação semelhante à noção de prestígio e estigma. As variantes conservadoras, tradicionais e já estáveis na comunidade de fala tendem a ser priorizadas em detrimento das variantes inovadoras, que, por vezes, sofrem restrição de uso ou estigma suficientes para impedir ou retardar a sua implementação ou estabilidade na comunidade de fala:

No nível morfossintático, um bom exemplo é o da formação do plural dos sintagmas nominais: apesar dos esforços da escola e de outras instituições no sentido de que todos os termos sujeitos à flexão recebam a marca /s/, cada vez mais a regra se simplifica. Se prestarmos bem atenção, notaremos que muitos falantes cultos já dizem meus amigo ou vinte pau, evitando a redundância própria da variante conservadora.” (MONTEIRO, 2000, p.66)

Essas avaliações e restrições impostas pelo falante são indispensáveis para a compreensão de como se dão os processos de mudança linguística.

3.1.1 Variação estável e mudança em curso

Sabemos que a língua não é estática nem homogênea: ela serve às sociedades, cujas necessidades existenciais – e, conseqüentemente, expressivas – mudam, sendo que a língua funciona como instrumento histórico da manifestação dessas necessidades, servindo aos

falantes através de movimentos de conservação e inovação. Os estudos variacionistas entendem a variação como pré-requisito para a mudança, sendo que, entretanto, essa relação não é dialógica: toda mudança é precedida por um período de variação, mas nem toda variação resulta em mudança. A mudança é inerente à natureza social da língua, e ela só se concretiza depois de um período de instabilidade, de um “ponto crítico do sistema e possibilidade de seleção entre dois modos equivalentes” (COSERIU, 1979, p. 104). Esse período pode ser superado, completando o processo propriamente dito da mudança, ou pode ser permanente. Chamamos esse período de instabilidade de variação linguística, ou, ainda mais especificamente, variação sociolinguística, já que tal instabilidade tem como agentes motivadores (ou limitadores) fatores de ordem social, além dos fatores linguísticos. Quando há um movimento de decisão em favor de uma ou outra variante, trata-se de um processo de mudança em curso.

Para visualizar a evolução de um processo de mudança diacrônica, seria ideal que se analisasse o comportamento da variante na comunidade em um período arbitrário, real de tempo. Esse estudo – denominado estudo da mudança em tempo real – pode ser realizado tanto através de textos antigos que registrem a variação do fenômeno, comparando-os com dados recentes, quanto realizando o mesmo estudo em dois períodos de tempo diferentes, na mesma comunidade e com os mesmos informantes.

É possível, também, verificar o comportamento das variantes fazendo inferências sobre o desenvolvimento, difusão e implementação de novas variantes em detrimento das que eram tradicionais na comunidade no momento em que o processo está em curso. Essa análise pode dar importantes indícios acerca do andamento de um processo de mudança e, ao contrário do que afirmavam os estruturalistas, é possível vislumbrar esse processo antes do resultado final, enquanto ele ainda não foi concluído. Esse é o estudo da mudança em tempo aparente, que consiste, então, em uma projeção sincrônica da mudança, que se dá no plano diacrônico. O estudo, nesse modelo, é realizado através da estratificação dos informantes em faixas etárias, e o comportamento das faixas revela os rumos da variável que está sendo analisada. São essas as bases teóricas, sobretudo no que diz respeito à concepção de língua, à noção de variável e à delimitação das variantes, à própria noção de variação e mudança, à estratificação dos informantes e à correlação entre as características sociais e linguísticas dos informantes que sustentarão a análise aqui proposta.

3.2 A CONSTITUIÇÃO DO *CORPUS*

É indispensável, na realização de um trabalho de cunho sociolinguístico, a atenção a critérios de levantamento e tratamento de dados estatísticos, já que as respostas numéricas são o subsídio para as conclusões alcançadas na pesquisa. A amostra de fala que servirá como *corpus* para a análise deve ser estratificada de acordo com as características linguísticas e extralinguísticas que serão consideradas, devendo esta amostra refletir, de forma satisfatória, a realidade sociolinguística da comunidade em análise. Posto isto, descreveremos, nessa seção, os passos seguidos para a constituição, tratamento e análise da amostra de fala que compõe esta pesquisa.

3.2.1 O Projeto Vertentes

Todo o acervo utilizado na presente pesquisa foi constituído no âmbito do Projeto Vertentes do Português popular da Bahia (www.vertentes.ufba.br) (doravante Projeto Vertentes), que é vinculado ao departamento de Letras Vernáculas da UFBA. O Projeto Vertentes visa a buscar, através da caracterização da realidade sociolinguística das regiões analisadas, os fundamentos empíricos para a formação da variante popular do português falado na Bahia, considerando que o contato entre línguas é o processo responsável pelas atuais características do PB. O *corpus* do Projeto Vertentes é composto por três acervos de fala: o português rural de comunidades afrobrasileiras isoladas, o português popular do interior do Estado da Bahia e O português popular da cidade de Salvador, sendo este último o *corpus* desta pesquisa, composto de entrevistas sociolinguísticas realizadas nos bairros de Liberdade, Cajazeiras, Itapuã e Subúrbio, que passamos a descrever nas seções seguintes.

3.2.2 As comunidades analisadas

A cidade de Salvador, uma das cidades mais importantes tanto da colônia quanto do Brasil independente, é, hoje, a terceira cidade mais populosa do Brasil. A transferência da capital para o Rio de Janeiro, o fim da comercialização de escravos ocasionada pelo encerramento da empresa escravocrata, a perda da hegemonia de produção e exportação de café, juntando-se ao fato de que boa parte da população soteropolitana era formada por ex-escravos com pouco ou nenhum poder de consumo, fez com que Salvador passasse, no início do século XX, por uma considerável estagnação econômica. Essa situação só foi revertida em meados do século XX, com a industrialização, o início das atividades da Petrobrás, a inauguração do Centro Industrial de Aratu e do Banco Nacional de Habitação, que atraiu

consideráveis fluxos migratórios para a capital, sendo que, em 1970, 70% da população soteropolitana era de origem migrante (SANTOS, 2008).

Como a língua está diretamente atrelada às transformações pelas quais as sociedades passam, os movimentos descritos acima tiveram importantes implicações sociolinguísticas. Temos, aí, dois movimentos populacionais distintos: Salvador passa de centro cujas atividades econômicas principais eram pautadas nas atividades portuárias e na exportação a cidade industrializada, tendo sua periferia majoritariamente ocupada por falantes advindos das áreas rurais do estado. Ao mesmo tempo, como é característico dos grandes centros urbanos do país, passa a ser, também, difusor dos padrões linguísticos dos grandes centros. (SANTOS, 2008).

É exatamente com vistas a investigar os padrões dessa fala popular urbana e buscando possíveis efeitos dos processos pretéritos de contato entre línguas que se justifica a necessidade de eleger Salvador como *corpus* de observação. Para tanto, era preciso eleger bairros da cidade que fossem representativos dessa diversidade, tanto no que diz respeito à localização geográfica quanto em relação à demografia e às datas de consolidação desses bairros, tarefa que foi realizada pela equipe do Projeto Vertentes.

Entretanto, uma grande dificuldade para a constituição do *corpus* de Salvador diz respeito ao fato de que, devido às políticas públicas de incentivo à educação básica, a maioria dos informantes teve escolarização mínima, sendo que não foram encontrados informantes da faixa etária I que fossem analfabetos. Além disso, alguns bairros têm data de fundação recente, o que impede a localização de informantes da faixa etária III que tenham nascido nas localidades. Nesses casos, decisões metodológicas, que serão descritas adiante, precisaram ser tomadas no intuito de adaptar a amostra às necessidades da pesquisa.

A seguir, falaremos um pouco sobre a socio-história dos bairros que serviram de *corpus* para a análise: Liberdade, Cajazeras, Itapuã e Subúrbio.

3.2.2.1 Liberdade

Situado na divisão entre a Cidade Baixa e a Cidade Alta de Salvador, Liberdade foi, por muito tempo, considerado o bairro com a maior população negra da cidade, perdendo esse posto, de acordo com o IBGE, em 2014. O bairro foi fundado no ano de 1823, nascendo a partir de uma estrada de terra que era utilizada para o transporte de gado, a Estrada das boiadas, que, após servir de caminho para os combatentes vitoriosos da independência da

Bahia, passou a ser chamada Estrada da Liberdade, o que explica o nome do bairro. Diversas localidades formam o bairro da Liberdade, como Soledade, Lapinha, Sieiro, Japão, Duque de Caxias, Curuzu, Cravinas, Bairro Guarani, Alegria, Jardim São Cristovão, São Lourenço e parte do Largo do Tanque e da Baixa do Fiscal. (BANDEIRA e MACAMBYRA, 2010).

A forte presença negra que caracteriza o bairro de Liberdade é devida à presença de escravos durante a constituição e estabelecimento do bairro. A atividade comercial é o ponto forte do bairro, mas boa parte dos moradores exerce atividades nos bairros vizinhos. Apesar de apresentar sérios problemas de violência e falta de infraestrutura, ocasionados pelo crescimento acelerado e não-planejado, o bairro tem forte apelo turístico, sediando os dois maiores blocos afro de Salvador: o Ilê Ayê e o Muzenza. (MACAMBYRA e BANDEIRA, 2010).

3.2.2.2 Cajazeiras

O bairro de Cajazeiras surgiu no início dos anos 80 a partir da desapropriação de quatro grandes fazendas (a Fazenda Jaguaripe de Cima, também conhecida como Fazenda Grande, a Fazenda Cajazeiras, a Fazenda Boa União e a Chácara Nogueira). É um dos bairros com maior contingente negro de Salvador, sendo formado por 9 setores, além dos bairros de Fazenda Grande 1, 2, 3 e 4, Águas Claras e Boca da Mata. Devido ao fato de ter sido planejado como solução para o crescimento desenfreado de Salvador, Cajazeiras passou a abrigar moradores oriundos do interior e de várias zonas da própria cidade de Salvador, o que fez com que, junto com as constantes invasões de encostas, fez com que o bairro alcançasse proporções territoriais muito grandes e passasse a apresentar problemas estruturais e sociais notáveis. Devido às suas grandes dimensões e à distância do centro de Salvador, Cajazeiras é um bairro que apresenta vida própria, contando com comércio, postos de saúde, bancos, supermercados, dentre outros serviços (TOURINHO E ANTONINO, 2010).

No que diz respeito à constituição do *corpus* de Cajazeiras, o Projeto Vertentes encontrou dificuldades em relação às faixas etárias, sendo adotada a seguinte estratégia metodológica para contornar o problema:

Por ser um bairro novo, que teve a sua expansão na década de 90, tornou-se praticamente impossível encontrar moradores de faixa II e III nascidos e criados no bairro; em decorrência disto, apenas os informantes de faixa I atenderam a esta característica, sendo os demais informantes do interior do estado, residindo no bairro há pelo menos 15 anos. (texto informado pelo site do Projeto Vertentes).

3.2.2.3 Itapuã

Situado na Orla Atlântica de Salvador, mais precisamente entre o bairro de Itapuã e a cidade de Lauro de Freitas, Itapuã foi, até os anos 50, uma colônia de pescadores. Sua escolha para compor o *corpus* do português popular da cidade de Salvador é explicada tanto pela localização do bairro, já que o mesmo se encontra na conexão entre Salvador e a região metropolitana, quanto pela sua força de representação da cultura da cidade, incluindo manifestações culturais como a tradicional Lavagem de Itapuã e o grupo Malê Debalê:

O grupo afro *Malê Debalê*, criado em 23 de março de 1979 por um grupo de moradores que almejavam ver o bairro de Itapuã representado no carnaval baiano, é um exemplo histórico representativo da forte identidade negra do bairro. Faz referência aos escravos africanos das etnias hauçá e nagô, de religião islâmica e responsáveis pela Revolta dos Malês em 1835, que escolheram o Abaeté como principal refúgio na cidade. (texto informado pelo site do Projeto Vertentes).

As entrevistas foram realizadas pela equipe do Projeto Vertentes com o intermédio de líderes locais, que, depois de serem contatados pelos entrevistadores, localizaram moradores que se encaixavam no perfil necessário.

3.2.2.4 Plataforma/Subúrbio

O quarto e último bairro incluído na amostra de fala da cidade de Salvador é o Subúrbio, o mais antigo bairro do subúrbio ferroviário, tendo surgido no final do século XIX. O perfil socioeconômico de Subúrbio assemelha-se aos dos demais bairros, já que essa localidade também sofre pelos problemas estruturais e sociais comuns aos bairros menos abastados de Salvador. Em relação à constituição do *corpus*, temos o seguinte:

No que diz respeito ao perfil dos moradores, um outro aspecto que chamou a atenção dos pesquisadores foi o clima de amizade e solidariedade entre eles, bem semelhante ao de comunidades do interior. É comum encontrar os vizinhos reunidos, no fim de tarde, nas portas das suas casas, para um "bate-papo" ou um joguinho de dominó. É natural que os filhos, ao se casarem, construam as suas casas bem próximas a dos seus pais. E é admirável o pacto de solidariedade que parece haver entre os moradores, uma vez que um vizinho está sempre disposto a ajudar o outro nos mais diversos aspectos. Esse perfil do bairro favoreceu amplamente a pesquisa na comunidade, pois o fato de conhecer dois ou três moradores possibilitava o acesso aos demais membros da comunidade e ajudava a dissolver o clima de desconfiança. Por outro lado, esse cenário típico de comunidades interioranas causou uma inquietação aos pesquisadores: a de saber se o apego à tradição do bairro e a sensação de pertencimento à comunidade influenciam ou refletem a manutenção da variedade local. (Texto informado pelo site do Projeto Vertentes).

3.2.3 O perfil dos informantes

A amostra analisada neste trabalho será composta por 48 informantes, 12 por bairro. Estes informantes foram estão estratificados de acordo com variáveis sociais específicas, que são: faixa etária (faixa I – 25 a 35 anos, faixa II – 45 a 55 anos e faixa III – mais de 65 anos), sexo, escolaridade (analfabeto ou semi-analfabeto) e estada fora da comunidade. Todos os informantes nasceram na localidade e ocupam a base da pirâmide social, com pouca ou nenhuma escolarização. Sendo assim, cada amostra foi composta por seis células, cada célula com 2 informantes, conforme o quadro abaixo:

Quadro 2 - A estratificação dos informantes

	Código do informante	Idade	Sexo		Código do informante	Idade	Sexo
	Liberdade	LIB- 1	29 anos		F	Itapuã	ITA-1
	LIB- 2	34 anos	M		ITA-2	35 anos	M
	LIB- 3	29 anos	F		ITA-3	30 anos	F
	LIB- 4	29 anos	M		ITA-4	28 anos	M
	LIB- 5	54 anos	F		ITA-5	52 anos	F
	LIB- 6	50 anos	M		ITA-6	55 anos	M
	LIB- 7	48 anos	F		ITA-7	45 anos	F
	LIB- 8	49 anos	M		ITA-8	49 anos	M
	LIB- 9	70 anos	F		ITA-9	89 anos	F
	LIB- 10	75 anos	M		ITA-10	94 anos	M
	LIB- 11	66 anos	F		ITA-11	81 anos	F
	LIB- 12	65 anos	M		ITA-12	67 anos	M
	Código do informante	Idade	Sexo		Código do informante	Idade	Sexo
	Subúrbio	SUB-1	33 anos		F	Cajazeiras	CAJ-1
	SUB-2	29 anos	M		CAJ-2	26 anos	M
	SUB-3	25 anos	F		CAJ-3	28 anos	F
	SUB-4	24 anos	M		CAJ-4	25 anos	M
	SUB-5	46 anos	F		CAJ-5	51 anos	F
	SUB-6	49 anos	M		CAJ-6	55 anos	M
	SUB-7	56 anos	F		CAJ-7	48 anos	F
	SUB-8	54 anos	M		CAJ-8	54 anos	M
	SUB-9	83 anos	F		CAJ-9	83 anos	F
	SUB-10	66 anos	M		CAJ-10	67 anos	M
	SUB-11	74 anos	F		CAJ-11	70 anos	F
	SUB-12	89 anos	M		CAJ-12	72 anos	M

3.2.4 As entrevistas

As entrevistas têm duração média de 50 minutos e seguem a orientação laboviana, em que se busca o vernáculo do informante através de uma interação que se assemelhe a uma conversa informal com o mínimo de interferência do entrevistador. Essa atitude visa a mitigar o que, na Sociolinguística Variacionista, é chamado de paradoxo do observador.

Sendo assim, os temas das conversas sempre estiveram voltados para questões relacionadas às comunidades, como história, costumes e questões políticas. Em nenhum momento da análise a identidade dos informantes é revelada, e todos declararam concordar com o uso dos dados fornecidos para fins científicos.

3.3 A VARIÁVEL DEPENDENTE

A variável dependente analisada aqui - os mecanismos de indeterminação do sujeito - compreende cinco variantes, que se mostraram produtivas em estudos realizados anteriormente e que representam cinco estratégias pronominais de indeterminação: *a gente*, *nós*, *você*, *eles* e a chamada categoria vazia (\emptyset + V3PS), formada pela terceira pessoa do singular sem referência anterior. Inicialmente, a variante formada pelo verbo na terceira pessoa do singular + o pronome *se* também estava incluída, mas, devido ao fato de não termos encontrado nenhuma ocorrência dessa variante no *corpus*, não pudemos analisá-la.

3.3.1 A gente

O pronome *a gente* passou a fazer parte do quadro pronominal brasileiro em variação com o pronome *nós*, derivando do sintagma nominal ‘a gente’, que significa agrupamento de pessoas, e sendo, atualmente, considerado como uma forma coloquial de pronome de primeira pessoa do plural. De acordo com a análise feita por Lopes (2003) sobre a inserção do *a gente* no quadro pronominal do português, este foi um processo lento, não isolado e encaixado na estrutura linguística e social, acontecendo em ritmos diferentes na fala e na escrita (LOPES, 2003, p.148):

A propagação ou emergência, a partir do século XV, das expressões nominais de tratamento, em substituição ao tratamento *vós*, a degradação semântica e a sua consolidação no século XVIII correspondem ao que foi observado na pronominalização de *a gente*. A perda de certas propriedades formais características do nome, como a subespecificação de número plural de *gente*, deu-se a partir do XVI. Além disso, o século XVIII marca a fase embrionária do seu processo de gramaticalização.

No PB, o pronome *a gente*, devido à sua generalidade, também é largamente utilizado como estratégia de indeterminação do sujeito, estando, no português popular, entre os recursos indeterminadores mais recorrentes, tal como nos exemplos a seguir:

(42) Eh! Eu gosto. Eu gosto muito disso. Tê folhas pa fazê chá, né, que *a gente...* na roça, *a gente* tem essas coisa. Hoje, *a gente* quase que já não tem mais isso. *A gente* num vê mais. Antigamente, eu entrava pelos mato, assim, tirava muita folha pa fazê xarope, 'que as pessoas me pedia. Hoje, *a gente* num acha mais folha pa fazê nada, raiz, nada. (CAJ 07)

3.3.2 Nós/V1pp

O pronome *nós* apresenta características semânticas que fazem com que ele seja largamente utilizado para se indeterminar o sujeito. Conforme Lopes (1998), sobre os pronomes pessoais plurais,

Ora, a noção de número implica o grupamento de elementos de mesma natureza e não é isso que ocorre com a forma *nós*, entendida como plural de *eu*, e *vós/vocês* como plural de *tu/você*. No primeiro caso, é inconcebível a junção de *eu+eu*, havendo, na verdade, várias possibilidades de compreensão: *eu+tu/você*, *eu+ele/ela*, *eu+vós/vocês*, *eu+eles*, *eu+todos*. No segundo caso, do mesmo modo, a forma plural refere-se a um conjunto de pessoas *com quem se fala*, admitindo também um valor indeterminado, abrangente, genérico e até difuso.

Sendo assim, o pronome *nós* (ou a forma verbal flexionada na primeira pessoa do plural) constitui um dos recursos de indeterminação do sujeito (43), embora, com a inclusão do *a gente* no quadro pronominal tenha feito com que essa estratégia tenha caído em relativo desuso, sendo mais frequente entre os falantes mais velhos, conforme veremos posteriormente, na análise.

(43) Mas eu me revolto com essa vida que *nós* tamo passano. Coisa triste. (LIB 11)

3.3.3 Você

A forma “vossa mercê” era, no século XV, destinada exclusivamente ao tratamento da nobreza portuguesa, sendo que, após modificações motivadas por alterações na estrutura social da sociedade portuguesa, chegou nos dias atuais na forma “você”. O “você”, apesar de ainda ser reconhecido pela maioria das gramáticas tradicionais como pronome de tratamento, sofreu alterações semânticas que possibilitaram sua inclusão no quadro de pronomes pessoais

do português, sendo, também, largamente utilizado como recurso de indeterminação do sujeito, com o podemos observar nos exemplos a seguir:

(44) Não tinha nada. Aqui *você* não tinha paladá de nada, não tinha nada pra *você* comê. (CAJ 09)

3.3.4 Eles/V3PP

Apesar de ser, por excelência, um pronome anafórico, já que retoma seres anteriormente mencionados no discurso, o pronome *eles*, quando utilizado em contexto de indeterminação do sujeito, perde essa função, passando a representar uma referência indeterminada. É preciso considerar, entretanto, que essa estratégia difere-se consideravelmente das estratégias de primeira e segunda, como *nós*, *você* e *a gente*, já que essas incluem o falante em suas referências (com exceção do *você*, que pode ou não incluir). O pronome *eles*, por sua vez, é a categoria da não-pessoa, o que faz com que seu sua abrangência, no que diz respeito à referencialidade, seja mais reduzida.

Para a nossa análise, consideramos tanto a ocorrência do pronome foneticamente realizado, como em (45), tanto a forma não realizada, com o verbo flexionado na terceira pessoa do plural (46).

(45) *Eles* viero e *fizero* o que *quisero* aqui. (ITA 02)

(46) É craro, é. Foi tão tanto que quando eu tive ele no IPERBA pa mim entregá deu trabalho. *Queriam* robá ele. (SUB 03)

3.3.5 Ø + V3PS

Essa é uma estratégia que, na maioria dos estudos variacionistas que focam o português popular brasileiro, apresenta maiores frequências, embora também seja utilizada por falantes com maior escolaridade. Lucchesi (2014), considera essa estratégia como diretamente relacionada ao processo de contato entre línguas na formação do PB, sendo, pois, fruto da perda do pronome indeterminador *se*, uma perda morfológica típica das línguas formadas em situações de contato (LUCCHESI, 2014, p.286):

Essa construção é bastante significativa, se pensarmos em termos de um possível efeito do contato entre línguas na formação do português brasileiro, em função de um nítido paralelo que pode ser traçado com pelo menos uma língua crioula de base lexical portuguesa, que também emprega essa estratégia de indeterminação do sujeito, divergindo do que ocorre no português de Portugal, que, em princípio, não aceitaria como gramatical tal construção. Dessa forma, essa construção pode ser vista como resultante de um processo histórico de simplificação morfológica, o

qual se eliminou o emprego do pronome se como índice de indeterminação do sujeito, de modo que o sujeito nulo sem referência anterior, por si só, já seria suficiente para expressar o sujeito de referência genérica.

(46) O grandão, *diz* que é do lá do Comércio. Ali, Pelôrinho. (ITA 09)

Para a análise da variável dependente, foram controladas oito variáveis linguísticas independentes e seis variáveis sociais independentes, que passamos a descrever nas seções seguintes.

3.4 AS VARIÁVEIS LINGUÍSTICAS INDEPENDENTES

Dentre todos os fatores linguísticos que podem influenciar na escolha de uma ou outra variante para a indeterminação do sujeito, analisaremos os seguintes, que se mostraram relevantes em estudos realizados anteriormente:

3.4.1 Realização fonética do sujeito indeterminado

Essa variável diz respeito à presença ou ausência de pronome sujeito na sentença. A análise dos resultados dessa variável nos permitirá verificar se há maior ou menor preferência, por parte dos falantes, por sujeitos foneticamente realizados para indeterminar o sujeito. Propomos essa variável para a análise porque ela pode fornecer dados importantes acerca do princípio “Evite pronome” (CHOMSKY, 1981), já que, de acordo com Duarte (1995), este princípio, apesar de dizer respeito à realização fonética dos sujeitos de referência definida, pode influenciar os sujeitos de referência arbitrária.

Essa variável inclui dois fatores:

1. Sujeito realizado foneticamente:

(47) Oxente, bastante. Cajazêra, antigamente, era o quê? Era um lugá... era pôco transporte, segurança era muito {precára}, banco *cê* num encontrava aqui em Cajazêra nenhum, agora *cê* já encontra quase todos os banco aqui. O que não tenho por aqui só é a Caixa Econômica, que não tem aqui, mas tem Bradesco, tem Banco do Brasil. (CAJ 08)

2. Sujeito não realizado foneticamente:

(48) Ø Dero, é, coisa de veneno, que Ø bota dento da, da carne, aí Ø mataro, que... dêxa lá, né? Eu num vô chegá pa dizê assim: "Cê matô meu gato?" Aí Ø vão querê ba... brigá, aí eu, aí eu num, num, eu sô nada contra isso aí, Ø matô, matô. (CAJ 02)

3.4.2 Desinência verbal

Pretendemos analisar, com essa variável, se há relação entre a redução do paradigma flexional dos verbos, fenômeno latente no português popular, e a escolha da estratégia de indeterminação por parte do falante. Para essa variável, consideraremos três possibilidades:

1. Terceira pessoa do singular

(49) Ø Faz trabalho, Ø bota coisa no meio da rua. Cês não encontraram não, no meio da rua assim? Galinha, faca...baraca, vejé, dinheiro... cê é doido! Alí é pra fazê o mal pra os outro. (LIB 01)

2. Primeira pessoa do plural

(50) *Nós* tamo viven... eu tô achando assim, *nós* tamo vivendo no tempo da ditadura ainda. (LIB 03)

3. Terceira pessoa do plural

(51) Hoje, *estão matano* por causa de {dloga}, hoje *tão matano* por causa de coisa, por causa de bestêra *tão matano* o ôto, então fica muito difícil, só Jesus mesmo pá tê misericórdia. (LIB 04)

3.4.3 Paralelismo formal

O paralelismo discursivo é um aspecto linguístico relevante para a maioria dos fenômenos linguísticos, seja qual for o idioma estudado. Largamente atestado nos estudos relacionados tanto à teoria da variação quanto à linguística funcionalista e à análise do discurso, o paralelismo é observado quando o uso de determinada variante induz a permanência da mesma variante no mesmo contexto comunicativo. De acordo com Scherre (2012, p. 31),

Esta restrição ou variável independente ocorre entre as cláusulas (plano discursivo), no interior da oração (plano oracional), no interior do sintagma (plano sintagmático), e entre palavras e no interior da palavra (plano da palavra).

Conforme dito acima, o paralelismo discursivo é observável em todos os níveis linguísticos. Entretanto, observamos, nessa análise, apenas o plano discursivo, já que

desejávamos verificar a ocorrência do fenômeno entre as orações. Sendo assim, observaremos se o paralelismo formal também é aplicável às estratégias de indeterminação, verificando se, ao escolher determinada estratégia, o falante tende a permanecer utilizando a mesma estratégia durante o turno de fala. Utilizamos, como critérios, a presença de sujeito com a mesma referência da ocorrência precedente, sendo que ambas não poderiam estar separadas pela eventual interrupção do documentador.

3.4.4 Referência ao falante

Analizamos, nessa variável, se o falante se inclui ou não no contexto determinado pela forma de sujeito utilizada. Visamos observar quais são as variáveis que tendem a ser favorecidas quando o falante se inclui na referência do sujeito indeterminado, acreditando que essas são as estratégias com referência mais abrangente. Consideramos, aqui, duas possibilidades:

1. O falante se inclui no contexto de fala:

(52) Fica toda machucada, e uma rôpa fica... se estraga, você ficá torceno, torceno, *você* fica polino a rôpa toda. Aí *você* esfrega ela, eu que... eu lavo pelo de... pelo a... pelo direito e pelo avesso que suja mais pelo avesso que *você* transpira muito aqui embáxo mesmo e aqui fica o sujo. Então *você* esfrega. (CAJ 08)

2. O falante não se inclui no contexto de fala:

(53) Sábado, sábado aqui, à noite, sábado e domingo, quando vocês vim uma vez assim de sábado pa tirá suas dúvida, que vocês venha com seus namorado ou seus, seus marido, chegue ali no largo. Eles, *eles* ficam bebeno, aí, *ligam* o som do carro, *abre* a mala do som do carro. Esse, esse que tava aqui comigo aqui é uma vítima disso aí. (SUB 05)

3.4.5 Tipo de verbo

Diz respeito ao caráter semântico do verbo da sentença em questão. Aqui, observaremos se há alguma relação entre o caráter semântico do verbo da oração e a estratégia indeterminadora utilizada pelo falante. Para essa análise, utilizamos a proposta de classificação semântica verbal apresentada por Mira Mateus et al (2003).

1. verbos transitivos *discendi*

(54) ...que era uma fêmea, né? Que é, me *falaro* que esse macho tem que tê três fêmea pa ele, eu não sabia. (ITA 07)

2. verbos transitivos cognitivos

(55) Eles *acha* que o menino tá indo lá se {'futá} pa dá dicas. (ITA 05)

3. verbos transitivos de ação pontual

(56) *Invadiro e roubaro* tudo. (CAJ 09)

4. verbos transitivos de ação não pontual

(57) Nós sempre *vai* lá pegá água. (SUB 10)

5. verbos de ligação/estativos

(58) Eles *era* tudo forte, era natural mesmo. (CAJ 12)

6. verbos intransitivos inacusativos

(59) Chegue lá agora, num hospital desse passano mal, cê *morre*. E esse dinhêro nosso vai pra onde? (LIB 08)

7. verbos intransitivos

(60) Você *anda* muito aqui. (LIB 06)

8. verbos de movimento

(61) A gente *ia* de ônibus ou bonde. (SUB 12)

9. verbos locativos

(62) A gente têm direito porque a gente *mora* aqui. (SUB 01)

10. verbos aspectuais

(63) Aí *começaro a colocá* barraquinha. (CAJ 09)

11. verbos modais

(64) A gente *tem que* ficá falano... (ITA 06)

12. verbos de suporte/leve

(65) Aí eles *deu uma melhorada*. (ITA 03)

13. verbos de posse

(66) Eles *têm* umas embarcação nova aí agora. (ITA 11)**3.4.6 Tipo de frase**

Essa variável pretende analisar se há diferença entre os recursos de indeterminação utilizados quando o falante está fazendo uma pergunta, uma afirmação ou quando está negando uma afirmação. Sendo assim, ela se apresenta em três possibilidades: frases interrogativas, negativas e afirmativas. Esperamos que a maioria das ocorrências seja de frases declarativas, já que esse tipo é favorecido pelo modo como as entrevistas são conduzidas.

1. Frases declarativas

(67) Nós brincava muito de noite. (SUB 06)

2. Frases negativas

(68) Agora tá perigoso, a gente num pode deixá as porta aberta. (ITA 06)

3. Frases interrogativas

(69) Se cê tivé um filho pa dá um socorro, cê num vai tê que saí? (SUB 04)

3.4.7 Nível de referencialidade do agente

A referencialidade consiste em um processo menos gramatical e mais pragmático, no qual o falante, durante o seu discurso, refere-se a uma entidade real ou imaginária, identificando tal entidade de forma mais ou menos clara. Analisar esse fenômeno numa perspectiva pragmática significa considerar que os níveis de referência só podem ser inferidos a partir da análise da fala. Além dessa interação, é preciso levar em consideração, também, a intenção de incluir ou excluir entidades desse contexto, conforme afirma Neves (2007). Para essa análise, consideraremos três níveis de indeterminação, que podem ser:

1. a *indeterminação universal*, que é verificada quando a referência do sujeito

indeterminado é retomável por qualquer entidade;

(70) Nós tem que acreditá que vai melhorá. (CAJ 06)

2. a *indeterminação parcial*, que ocorre quando a referência o sujeito é parcialmente referencial;

(71) Aqui você vai no centro com 50 reais, num traz nada. (LIB 08)

3. o *grupo específico não definido*, que diz respeito à referência a um grupo menor, ou a apenas um indivíduo.

(72) Tocar fogo em tudo lá. (SUB 02)

3.4.8 Modo

Observaremos, aqui, se os modos verbais condicionam a escolha das estratégias de indeterminação. Analisaremos, também, se há a prevalência de algum dos modos verbais nos contextos de indeterminação. Essa variável refere-se à diferenciação semântica do verbo da oração em questão, que pode ser classificado como estando em dois modos:

1. Realis;

(73) Por causa de bestêra tão matano os outros. (CAJ 09)

2. Irrealis.

(74) Aqui se você tivesse uma vizinha que você considerasse, você fazia de tudo por ela. (ITA 04)

3.5 AS VARIÁVEIS SOCIAIS

Os fatores sociais essenciais para os pressupostos teórico-metodológicos dos estudos sociolinguísticos, já que os mesmos compreendem que a variação e as mudanças inerentes às línguas humanas condicionadas não apenas por questões internas, mas também por fatores externos a ela.

Como já dissemos anteriormente, a língua, por ser produto e meio de interação, social não deve ser estudada separadamente das questões sociais que a influenciam. Nessa análise, foram consideradas as variáveis sociais tradicionalmente consideradas nos estudos da Sociolinguística quantitativa. São elas:

3.5.1 Sexo

Essa é uma variável social bastante importante para os estudos sociolinguísticos. Embora as explicações e hipóteses clássicas acerca dessa variável girem em torno das atitudes mais conservadoras das mulheres em relação aos homens, devido à falta de igualdade entre os gêneros e a necessidade de resistir às formas inovadoras (CHAMBERS & TRUDGIL, 1994), é fundamental levar em consideração a comunidade de fala individualmente e o fenômeno analisado. Não é nunca uma questão genética que determina se mulheres ou homens terão determinado comportamento linguístico ou social, e sim o contexto sociolinguístico e os papéis que exercem na comunidade de fala em que vivem.

As hipóteses mais recorrentes dos estudos sociolinguísticos em relação ao sexo afirmam que as mulheres costumam apresentar comportamento mais conservador quando o movimento de mudança envolve variantes estigmatizadas. Quando há estigma, a tendência das mulheres é de utilizar as variantes inovadoras. Acreditamos que a indeterminação do sujeito não é um fenômeno que envolve estigma, o que nos leva a crer que as mulheres darão preferência às variantes mais inovadoras.

3.5.2 Faixa etária

A faixa etária é uma das mais tradicionais variáveis sociais dos estudos no âmbito da sociolinguística variacionista, sendo que essa variável é incluída nesses estudos com vistas a incluir no escopo da análise uma amostra de indivíduos que representem satisfatoriamente a comunidade de fala em questão. A hipótese principal sobre essa variável é que gerações diferentes refletem estágios diferentes da língua, sendo que a tendência é que os falantes mais jovens introduzam a variante inovadora, enquanto as faixas etárias mais avançadas tendem a utilizar as variantes mais inovadoras. Segundo Paiva & Duarte (2003, p.14, apud ARAÚJO, 2007):

Um dos princípios mais claros da Teoria da Variação é o de que as línguas naturais estão em constante variação. Assim, a Sociolinguística Variacionista postula que as mudanças possam ser apreendidas no seu curso de implementação através do que se denominou análise em tempo aparente. Vários estudos sociolinguísticos levam em conta essa análise, visto que, como sugerem muitos pesquisadores “o estudo da mudança em tempo aparente está baseado no pressuposto de que diferenças linguísticas entre gerações podem espelhar desenvolvimentos diacrônicos, quando outros fatores se mantêm constantes.

A análise realizada aqui será feita em tempo aparente, e consideraremos três faixas etárias:

Quadro 3 - Faixas etárias

FAIXA I	25-35 anos
FAIXA II	45-55 anos
FAIXA III	Mais de 65 anos

A distribuição das faixas etárias poderá nos mostrar quais as variantes que estão entrando na comunidade de fala, sendo mais frequentes na fala das faixas etárias mais novas, e quais estão caindo em desuso, mais frequentes na faixa mais velha.

3.5.3 Estada fora da comunidade

Com essa variável, intenta-se verificar a influência de outras comunidades de fala para a variedade de língua que está sendo analisada. Sendo assim, registraremos se há diferença entre as variantes utilizadas por falantes que nunca permaneceram mais que seis meses fora da comunidade, em contra posição aos falantes que já moraram em outras comunidades por mais de seis meses.

3.5.4 Escolaridade

A variável escolaridade também é recorrente nos estudos sociolinguísticos, estando diretamente relacionada à classe social à qual o indivíduo pertence. A hipótese principal acerca dessa variável é que a quantidade de anos de escolarização do indivíduo influencia diretamente no seu repertório linguístico. Na análise aqui realizada, por se tratar do português popular, consideramos apenas informantes analfabetos ou semi-analfabetos. É importante ressaltar, entretanto, que, devido ao fato de a análise ter sido realizada em uma capital, não foi possível realizar a estratificação proporcional entre informantes analfabetos e semi-analfabetos, visto que a maioria da população já teve acesso a alguma escolarização, ainda que por um curto período.

3.5.5 Localidade

A análise das diferenças entre localidades busca evidenciar as peculiaridades de cada espaço geográfico, considerando a diversidade cultural dos falantes, a história da formação daquela comunidade, as relações linguísticas e das interferências das outras línguas que se fizeram presente nesse espaço geográfico ao longo do tempo. Essas informações são importantes para confrontar as diferentes variedades usadas em espaços distintos, e, ao serem confrontadas, podem conduzir aos estudos dialetais a uma possível

resposta sobre os fenômenos que permeiam a variação linguística espacial (CARDOSO, 2010, p. 15).

Esta análise considerará quatro localidades, sendo estes bairros da cidade de Salvador: Liberdade, Cajazeiras, Itapuã e Subúrbio.

3.6 CONSTITUIÇÃO DOS DADOS E SUPORTE QUANTITATIVO

Depois de gravadas, as entrevistas foram transcritas e todas as ocorrências de sujeito indeterminado foram identificadas e codificadas de acordo com uma chave de codificação predefinida, que contém todas as variáveis linguísticas e sociais descritas acima. Cada variável independente recebeu um código, de acordo com o contexto discursivo da ocorrência e das características sociais do falante. Depois de codificadas, as ocorrências foram submetidas ao pacote de programas estatísticos GOLDVARB X (SANKOFF; TAGLIAMONTE; SMITH, 2005) que calcula as frequências de cada variante que compõe a variável dependente, identificando quais fatores linguísticos e sociais condicionam a variável e apresentando os seus respectivos pesos relativos.

O uso do método multivariado na análise linguística justifica-se pela natureza desse tipo de problema/fenômeno. A análise de fenômenos linguísticos revela a atuação concomitante de vários fatores de ordem linguística e extralinguística. Dessa forma, uma análise univariada, que trate separadamente essas influências, não comporta a complexidade e ação em cadeia dos fatores sobre o fenômeno em questão. A adoção de uma análise multivariada, por sua vez, é “mensurar” a influência mútua de fatores e “dará resultados mais precisos, porque, ao mesmo tempo em que computa o efeito de uma variável independente, ela controla explicitamente o efeito de todas as outras variáveis independentes conhecidas” (GUY; ZILLES, 2007, p.34)

Vimos, aqui, a fundamentação teórica na qual o trabalho se baseia: a Sociolinguística Variacionista. Descrevemos detalhadamente a amostra de fala utilizada e os critérios metodológicos que seguimos para a análise. Descrevemos, também, a variável dependente e as variáveis linguísticas e sociais cujas hipóteses serão testadas, visando verificar a influência dessas variáveis para o fenômeno.

4 ANÁLISE DOS DADOS

Neste capítulo, descreveremos os resultados encontrados na análise das cinco estratégias indeterminadoras referidas: *a gente*, *nós*, *eles*, *você* e $\emptyset + V3PS$. Para tanto, confrontamos cada uma das estratégias com as demais, e apresentaremos, para cada uma dessas estratégias, as variáveis linguísticas e sociais consideradas estatisticamente relevantes.

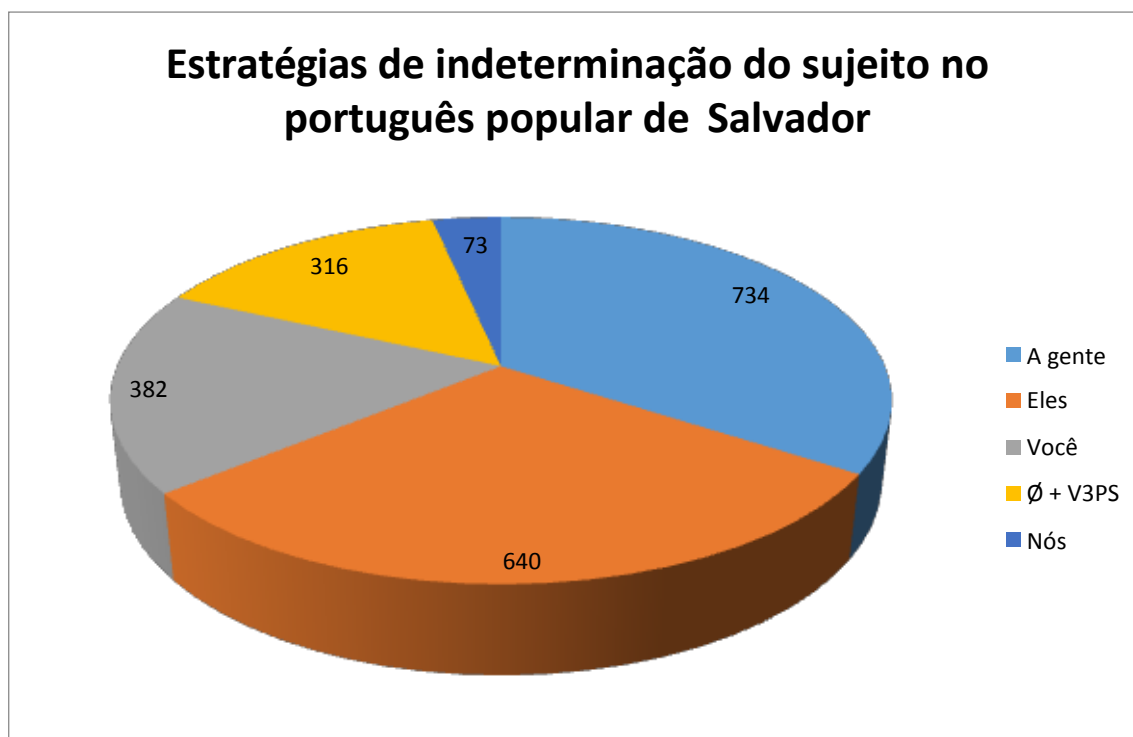
4.1 DISTRIBUIÇÃO GERAL DA VARIÁVEL DEPENDENTE

Para a análise quantitativa foram codificadas 2.145 ocorrências de sujeito indeterminado. A estratégia de indeterminação que se apresentou em maior número foi a com o uso do pronome *a gente*, totalizando 734 ocorrências, o que corresponde a 34,2% do total. Em seguida, a estratégia utilizada com mais frequência foi o pronome *eles*, com 640 ocorrências (29,8% do total). A terceira estratégia mais utilizada foi o pronome *você*, totalizando 17,8% do total, ou 382 ocorrências. A estratégia que aparece como a quarta mais utilizada foi a forma $\emptyset + V3PS$, totalizando 14,7%, ou 316 ocorrências. Em quinto e último lugar, temos o uso do pronome *nós*, com frequência de uso de 3,4%, o que corresponde a 73 ocorrências. No *corpus* analisado, não há nenhuma ocorrência de uso do pronome *se* como estratégia de indeterminação do sujeito. A distribuição geral das ocorrências é apresentada na tabela a seguir:

Tabela 15 - Frequência geral das estratégias de indeterminação do sujeito no português popular de Salvador

Estratégia de indeterminação	Número de ocorrências	Percentual
A gente	734	34,2%
Eles	640	29,8%
Você	382	17,8%
$\emptyset + V3PS$	316	14,7%
Nós	73	3,4%
Total	2145	100%

Gráfico 3 - Estratégias de indeterminação do sujeito no português popular de Salvador



Após apresentar um panorama geral da frequência de uso dessas estratégias, observamos a forma como as mesmas estão distribuídas no que diz respeito à realização/não realização fonética dos sujeitos indeterminados. Conforme já mencionamos em capítulos anteriores, Duarte (1995, 2007, 2012), com base em uma análise variacionista articulada com pressupostos gerativistas relativos ao Parâmetro do Sujeito Nulo, defende a hipótese de que o português brasileiro estaria deixando de ser uma língua “evite pronome” (como é o caso do italiano, língua de sujeito nulo prototípica no grupo românico), já que a falta de especificação da pessoa do discurso através da morfologia flexional induz o falante a realizar o sujeito com mais frequência. Essa alteração também teria atingido os sujeitos de referência genérica, uma importante evidência do “encaixamento” em decorrência dessa mudança: as ocorrências de sujeito definido expreso levariam à realização mais frequente de formas pronominais com valor indeterminador. Para verificar a relevância dessa hipótese no português popular, analisamos a frequência de realização fonética dos sujeitos indeterminados. Os resultados obtidos estão apresentados na tabela a seguir:

Tabela 16 - A indeterminação do sujeito de acordo com a variável Realização fonética do sujeito indeterminado

Estratégia de indeterminação	Realização fonética/Total	Percentual
A gente	568/734	77%
Eles	182/640	28%
Você	310/382	81%
Ø + V3PS	--	--
Nós	59/73	80%
Total	1119/2145	52,2%

Os resultados acima sugerem a uma possível correlação entre a realização fonética dos sujeitos indeterminados e a perda do princípio “Evite pronome” do português brasileiro (CHOMSKY, 1981). Observamos que, das 734 ocorrências com o pronome *a gente*, 568, ou 77% do total, tiveram o pronome foneticamente realizado. Situação semelhante ocorre com os pronomes *você* e *nós*, que apresentaram frequência de realização fonética de 81% e 80%, respectivamente. A estratégia Ø + V3PS, por ser caracterizada pelo lugar vazio do pronome sujeito, não apresentou ocorrências de sujeitos foneticamente realizados. O caso do pronome *eles*, apesar de, aparentemente, se opor à hipótese inicial, relacionada à perda do princípio “evite pronome” é, na realidade, um fator confirmatório da hipótese.

Duarte (1995, 2007, 2012), afirma que a inserção de novos pronomes no sistema pronominal brasileiro fez com que a morfologia flexional dos verbos mudasse no PB, o que altera consideravelmente as regras de concordância verbal e nominal. No que diz respeito à concordância verbal, temos, atualmente, um paradigma verbal que assume apenas três formas. Portanto, há, no português popular, o sincretismo das flexões verbais de quase todas as pessoas do discurso, sendo que apenas a primeira pessoa do singular não sofre esse tipo de variação:

Quadro 4 - Paradigma verbal

<p>Eu como Tu/Você come Ele/Ela come Nós/A gente come(mos) Vocês come(m) Eles/Elas come(m)</p>

Podemos perceber esse sincretismo da morfologia verbal ao analisar a distribuição das formas de indeterminação verbal entre as morfologias verbais, conforme descrito na tabela abaixo:

Tabela 17- A indeterminação do sujeito de acordo com a variável Desinência verbal

Desinência verbal	A gente		Eles		Você		Ø + V3PS		N		Total/ %
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Fre	%	
Terceira pessoa do singular	732	45,2	162	10	382	23,6	316	19,5	28	1,7	1620/76,1
Primeira pessoa do plural	2	4,3	--	--	--	--	--	--	45	95,7	47/2,2
Terceira pessoa do plural	--	--	462	100	--	--	--	--	--	-	462/21,7
Total	734	34,5	624	29,3	382	17,9	316	14,8	73	3,4	2129

Observamos, de acordo com a tabela acima, que a forma da terceira pessoa do singular é largamente utilizada para todos os pronomes, estando presente em 76,1% das ocorrências. A primeira e a terceira pessoa do plural se apresentam em números muito menores (2,2% e 21,7%, respectivamente), o que nos mostra que a morfologia flexional está realmente reduzida no português popular. Entretanto, o caso do pronome *eles* se delinea de forma diferente. Percebemos que a maior parte das ocorrências dessa forma pronominal é utilizada com verbos na terceira pessoa do plural, conforme os exemplos a seguir:

(75) Do nada, assim, eu passano, *eles* me dero um murro, assim, eu caí. [No caí], *eles* começaro me furá. (ITA 04)

(76) Foi, me *eles* assaltaro, *eles* levaro... levaro o celulá. (LIB 04)

Sendo assim, percebemos que o pronome *eles*, quando utilizado como estratégia de indeterminação do sujeito, se apresenta com morfologia flexional forte, o que faz com que a realização fonética do sujeito se torne menos necessária e, conseqüentemente, menos frequente.

Outra variável que analisamos aqui é o Tipo de verbo, com vistas a esclarecer se o caráter semântico dos verbos presentes nas orações influencia na escolha de uma ou outra estratégia de indeterminação. Os resultados obtidos nessa variável estão descritos na tabela a seguir:

Tabela 18 - A indeterminação do sujeito e a variável Tipo de verbo

Tipo de verbo	A gente		Eles		Você		Ø + V3PS		Nós		Total/%
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	
Discendi	11	12,5	29	33	9	10,2	37	42	2	2,3	88/4,1
Cognitivos	32	29,9	29	27,1	34	31,8	8	7,5	4	3,7	107/5
Ação	410	31,4	491	37,6	192	14,7	181	13,9	32	2,5	1306/60,9
Ligação	47	38,2	26	21,1	28	22,8	7	5,7	15	12,2	123/5,7
Inacusativos	4	57,1	1	14,3	2	28,6	--	--	--	--	7/0,3
Intransitivos	55	54,5	15	14,9	20	19,8	7	6,9	4	4	101/4,7
Movimento	57	47,1	22	18,2	23	19	14	11,6	5	4,1	121/5,6
Locativos	2	100	--	--	--	--	--	--	--	--	2/0,1
Aspectuais	4	28,6	9	64,3	1	7,1	--	--	--	--	14/0,7
Modais	74	37	7	35,5	58	29	58	29	3	1,5	200/9,3
De suporte/leve	4	28,6	8	57,1	1	7,1	1	7,1	--	--	14/0,7
De posse	34	54,8	3	4,8	14	22,6	3	4,8	8	12,9	62/2,9
Total	734	34,2	640	29,8	382	17,8	316	14,7	73	3,4	2145

Os resultados nos mostram que, como já se esperava, o tipo de verbo encontrado em maior número no *corpus* são os verbos de ação, estando presentes em mais da metade das

ocorrências (60,9%). Em seguida, temos os verbos modais, com 9,3%, de ligação, com 5,7%, de movimento, com 5,6%, e os intransitivos, que totalizaram 4,7%. Os demais tipos verbais não apresentaram frequências significativas, e as estratégias de indeterminação se encontram distribuídas de modo equilibrado entre as estratégias, o que nos leva a afirmar que essa não é uma variável relevante para a escolha de uma ou outra forma pronominal de indeterminação.

4.2 NÍVEIS DE REFERENCIALIDADE DO AGENTE

Consideraremos, nessa análise, três níveis de gradação para a indeterminação do sujeito, já que se sabe que este não é um fenômeno uniforme, mas sim observável em níveis de gradação diferenciados. Os graus considerados foram a *indeterminação universal*, que é verificada quando a referência do sujeito indeterminável é ocupável por qualquer ser potencialmente retomável pela expressão utilizada pelo falante, ou, de acordo com Dik apud Pereira (2003, p.477), por “qualquer entidade que satisfaça as restrições de seleção impostas para esta posição argumental”, a *indeterminação parcial*, que ocorre quando a referência o sujeito é parcialmente referencial, já que inclui indivíduos de um grupo mais específico (geralmente, os habitantes da localidade que o falante reside), e o *grupo específico não definido*, verificado quando a referência é um grupo menor, ou até mesmo um só indivíduo, cuja identidade não pode ser retomada anaforicamente. É válido considerar, entretanto, que o estabelecimento da referencialidade, no discurso, não é aleatória, sendo motivada por questões puramente contextuais/pragmáticas, sendo que tal decisão está diretamente relacionada com a intencionalidade do falante.

A distribuição das ocorrências de acordo com os diferentes níveis de referencialidade está apresentada na tabela 19:

Tabela 19 - A indeterminação do sujeito de acordo com a variável Nível de referencialidade do sujeito

Nível de referencialidade	A gente		Eles		Você		Ø + V3PS		Nós		Total/%
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	
Genérico	99	48,8	--	--	68	33,5	29	14,3	7	3,4	203/9,5
Parcial	551	46,8	97	8,2	306	26	165	14	59	5	1178/54,9
Específico	84	11	543	71,1	8	1	122	16	7	0,9	764/35,6
Total	734	34,2	640	29,8	382	17,8	316	14,7	73	3,4	2145

4.2.1 Nível genérico

De acordo com os dados apresentados na tabela, podemos observar que o pronome *a gente* é o mais recorrente em contextos em que a indeterminação é genérica, ou seja, em contextos em que o sujeito indeterminado pode ser preenchido por qualquer entidade. Para esse nível, obteve-se 48,8% das ocorrências de *a gente*, seguidas por 33,5% de ocorrências com o pronome *você*, 14,6% da forma Ø + V3PS e 3,4% do pronome *nós*. Essa é o nível em que há maior equilíbrio entre as estratégias de indeterminação e as suas frequências. Podemos observar o comportamento dessas variantes em relação a esse nível nos exemplos a seguir.

(77) *A gente* num deve desacreditá em nada! (ITA 01)

(78) Quando *a gente* morre, 'sa menina, num leva nada. (ITA 11)

(79) Porque quando *você* mora num lugá atribulado, **você** não tem tempo nem pra si próprio. (CAJ 03)

(80) Que num tem coisa melhô do que a pessoa ajudá o próximo. Hoje *você* tá, tá no auge, lá embáxo. De repente *cê* tá no auge cá em cima. (LIB 08)

(81) Ø Tem que se virá, né? Num tem jeito, Ø tem que corrê atrás. (SUB 02)

(82) Eu vô fazê o quê? Que o mundo é esse que *nós* estamo vivo! (LIB 01)

Não foram observadas ocorrências do pronome *eles* para esse nível de indeterminação, o que nos leva a inferir que essa estratégia é mais comumente utilizada em contextos referenciais mais restritos, conforme veremos posteriormente.

4.2.2 Grupo parcialmente definido

Apesar de todas as estratégias de indeterminação apresentarem ocorrências nesse nível de indeterminação, podemos observar que há uma prevalência do pronome *a gente*, com 46,8% das ocorrências. Em seguida, temos o pronome *você*, com 26%, seguido da forma \emptyset + V3PS, com 14% das ocorrências, do pronome *eles*, com 8,2%. Por último, temos o pronome *nós*, totalizando 5% das ocorrências.

(83) Foi agora que... eu tava até comentano com o professô, porque ele é um professô ótimo, *a gente* pode se abrí, ele conversa com *a gente* assim... (LIB 01)

(84) É. Tem em qualqué lugá a violência, mas aqui, 'té agora ainda num aconteceu nada, aqui nesse pedaço onde *a gente* mora. (LIB 05)

(85) Aqui era tudo bom, era mais fácil, tudo que *você* comprava, era melhó de que hoje em dia, que *você* vai hoje vai com cento e cinquenta reais, *você* num traz nada. (LIB 05)

(86) \emptyset fazia tudo, hoje já num \emptyset faz mais, já deu muito bom aí o point. (CAJ 12)

(87) \emptyset *Dissero* que eles não ia fazer nada por nós daqui. (CAJ 09)

(88) *Nós* daqui tamo vivo no dias difícil. (ITA 04)

4.2.3 Nível específico

As ocorrências no nível específico de indeterminação são majoritariamente favorecedoras do uso do pronome *eles*, o que pode ser explicado pela maior especificidade desse pronome em relação à sua abrangência, já que o mesmo se restringe apenas ao universo das terceiras pessoas do plural, que é um nível com abrangência mais específica. Para este

pronome, temos um total de 71,1% das ocorrências. Em seguida, temos a forma Ø + V3PS, utilizada em um total de 16% das ocorrências. Os pronomes *a gente*, *você* e *nós* apresentaram, para este nível de indeterminação, frequências de 11%, 1% e 0,9%, respectivamente.

(89) *Eles* pegaro o rapaz, *eles* batero, Ø fizeram de tudo com ele. (CAJ 12)

(90) Me Ø assaltaro ali mês passado. (ITA 10)

(91) Ø Tá fazendo um posto médico lá, Ø tá construindo. (SUB 01)

(92) *A gente* já visitou muita gente durante as missões. (CAJ 09)

(93) Quando *cê* vai vendê o pêscoço, eles não querem pagá. (CAJ 03)

(94) *Nós* andou aquilo tudo ali, depois *nós* veio pra o lado de cá. (SUB 09)

A seguir, passaremos a comentar e comparar as análises por GOLDVARB-X do efeito das variáveis independentes linguísticas e extralinguísticas sobre cada variante da variável dependente.

4.3 A INDETERMINAÇÃO DO SUJEITO NO PORTUGUÊS POPULAR DE SALVADOR

4.3.1 *A gente*

A estratégia utilizada com maior frequência foi a forma *a gente*, que esteve presente em 734 ocorrências, 34,2% do total. Para a submissão das ocorrências ao programa GOLDVARB X, precisamos excluir a variável explanatória inclui/não inclui o falante, visto que todas as ocorrências com essa forma pronominal pressupõem a inclusão do falante no contexto de fala, o que faz com que o programa estatístico apresente *knock-outs* nessa variável. Após submeter os dados à análise, o programa considerou como estatisticamente relevantes as seguintes variáveis, selecionadas nessa ordem de importância:

Tabela 20 - Variáveis selecionadas – *a gente*

- 1º Nível de referencialidade do agente
- 2º Realização fonética do sujeito indeterminado
- 3º Localidade
- 4º Modo
- 5º Sexo

A primeira variável considerada estatisticamente relevante para essa estratégia é Nível de referencialidade do agente. Os resultados obtidos para essa variável estão apresentados na tabela a seguir:

Tabela 21- O pronome *a gente* segundo a variável Nível de referencialidade do agente

Nível de referencialidade do agente	Número de ocorrências/Total	Percentual	Peso relativo
Genérico	99/203	48,8%	.681
Parcial	551/1178	46,8%	.653
Indefinido	84/764	11%	.235
Total	734/2145	34,2%	--

Nível de significância: .046

O pronome *a gente* é mais utilizado para os dois níveis de indeterminação mais abrangentes, o nível genérico e o nível parcial, apresentando, nesses níveis, respectivamente, .681 e .653. Com isso, podemos perceber que as estratégias de primeira pessoa do plural sempre envolvem contextos mais abrangentes, que incluem o próprio falante no contexto de fala.

A segunda variável selecionada como estatisticamente relevante é Realização fonética do sujeito indeterminado:

Tabela 22 - O pronome *a gente* segundo a variável Realização fonética do sujeito indeterminado

Realização fonética do sujeito indeterminado	Número de ocorrências/Total	Percentual	Peso relativo
Sujeito realizado	568/1119	50,8%	.667
Sujeito não realizado	166/1026	16,2%	.320
Total	734/2145	34,2%	--

Nível de significância:.046

A tabela acima mostra que o pronome *a gente* é majoritariamente utilizado com o pronome foneticamente realizado, sendo que ocorrências apontam peso relativo de .667.

Como já foi comentado anteriormente, as formas pronominais realizadas foneticamente são mais frequentes para quase todas as estratégias de indeterminação. A seguir, os resultados obtidos para a variável social Localidade do informante:

Tabela 23 - O pronome *a gente* segundo a variável Localidade

Localidade	Número de ocorrências/Total	Percentual	Peso relativo
Liberdade	128/509	25,1%	.306
Cajazeiras	171/495	34,5%	.811
Itapuã	211/627	33,7%	.361
Subúrbio	224/514	43,6%	.541
Total	734/2145	34,2%	--

Nível de significância:.046

O uso da estratégia *a gente* foi favorecido pelo bairro de Cajazeiras, com peso relativo de .811. Subúrbio, Itapuã e Liberdade apresentaram, respectivamente, .541, .361 e .306 de peso relativo.

Os resultados a seguir referem-se à quarta variável selecionada como estatisticamente relevante para a estratégia *a gente*: o Modo verbal.

Tabela 24 - O pronome *a gente* segundo a variável Modo verbal

Modo	Número de ocorrências/Total	Percentual	Peso relativo
Realis	721/2077	34,7%	.511
Irrealis	13/68	19,1%	.217
Total	734/2145	34,2%	--

Nível de significância:.046

Há um favorecimento discreto do uso da forma *a gente* em orações em que o verbo se encontra no modo Realis, equivalente ao que a tradição gramatical reconhece como modo indicativo. Nesse modo, a estratégia *a gente* apresenta peso relativo de .511. Para o modo Irrealis, correspondente ao modo subjuntivo, o peso relativo é .217. Os dois modos estão

apresentados nos exemplos abaixo, sendo o exemplo (96) correspondente ao modo Realis e o exemplo (97) correspondente ao modo Irrealis:

(95) *Se a gente* quisé uma folha pa fazê chá, tem que ir lá no São Joaquim comprá. (LIB 11)

(97) *A gente* fi... dormia de porta aberta. *A gente* queria comê, *a gente* arranjava qualqué coisa assim pa comê. (CAJ 08)

A variável Sexo foi a última selecionada pelo programa para a estratégia *a gente*.

Tabela 25 - O pronome *a gente* segundo a variável Sexo

Sexo	Número de ocorrências/Total	Percentual	Peso relativo
Feminino	343/1025	33,5%	.459
Masculino	391/1120	34,9%	.537
Total	734/2145	34,2%	--

Nível de significância:.046

Os resultados para essa variável foram muito próximos, sendo que a estratégia *a gente* é levemente favorecida pelos homens, com peso relativo de .537, e levemente desfavorecida pelas mulheres, com um peso relativo de .459.

4.3.2 Eles

A estratégia com a segunda maior frequência foi o pronome *eles*, estando presente em 640 ocorrências, o que corresponde a 29,8% do total. Para a análise dessa estratégia, também desconsideramos a variável Inclui/não inclui o falante, já que não existem ocorrências com este pronome que incluam o falante no contexto de fala, o que leva o programa a apresentar *knock-outs* na rodada. As variáveis selecionadas pelo programa como estatisticamente relevantes foram as seguintes, em ordem de importância:

Tabela 26 - Variáveis selecionadas – *eles*

- 1º Realização fonética do sujeito indeterminado
- 2º Sexo
- 3º Modo
- 4º Tipo de frase
- 5º Localidade

Os resultados da primeira variável selecionada como estatisticamente relevante seguem na tabela abaixo:

Tabela 27 - O pronome *eles* segundo a variável Realização fonética do sujeito indeterminado

Realização fonética do sujeito indeterminado	Número de ocorrências/Total	Percentual	Peso relativo
Sujeito realizado	182/1119	16,3%	.344
Sujeito não realizado	458/1026	44,6%	.669
Total	640/2145	29,8%	--

Nível de significância: .017

Ao contrário das demais estratégias, podemos perceber que em contextos de uso do pronome *eles*, é mais frequente a não realização do sujeito, sendo que, nesse contexto, o peso relativo é de .669. Nesses contextos, conforme já dissemos anteriormente, é possível que se associe à morfologia verbal da terceira pessoa do plural a alta frequência de sujeitos não realizados foneticamente.

A seguir, os resultados obtidos para a variável Sexo:

Tabela 28 - O pronome *eles* segundo a variável Sexo do informante

Sexo	Número de ocorrências/Total	Percentual	Peso relativo
Feminino	349/1025	34%	.559
Masculino	291/1120	26%	.446
Total	640/2145	29,8%	--

Nível de significância: .017

Como mostram os números acima, a estratégia *eles* é levemente favorecida pelas mulheres, com peso relativo de .559, enquanto que, para os homens, é levemente desfavorecida, com um peso relativo de .446.

Os resultados abaixo dizem respeito à variável Modo verbal.

Tabela 29 - O pronome *eles* segundo a variável Modo verbal

Modo verbal	Número de ocorrências/Total	Percentual	Peso relativo
Realis	632/2077	30,4%	.509
Irrealis	8/68	11,8%	.261
Total	640/2145	29,8%	--

Nível de significância: .017

O modo Realis, ou seja, o modo verbal indicativo, favorece a estratégia indeterminadora *eles*, com peso relativo .509. Em contrapartida, a forma Irrealis, correspondente ao modo subjuntivo, é bastante desfavorável para essa estratégia, apresentando um peso relativo de .261.

A penúltima variável selecionada foi Tipo de frase:

Tabela 30 - O pronome *eles* segundo a variável Tipo de frase

Tipo de frase	Número de ocorrências/Total	Percentual	Peso relativo
Declarativa	598/1906	31,4%	.513
Negativa	42/239	17,6%	.400
Interrogativa	--	--	--
Total	640/2145	29,8%	--

Nível de significância: .017

De acordo com os resultados apresentados, o pronome *eles* é favorecido pelas frases declarativas, com peso relativo de .513 (98). Com frases negativas, o peso relativo é levemente desfavorável, caindo para .400 (99).

(98) Eu levei uma caixa cheia, *eles* só dero doze reais. A vida de catá papel é trinta centavos o... eu já catei papelão na rua, eu já catei... catá essas garrafa peti que tem aí. Cê sabe quanto foi que deu? Fui com sacão cheio, cheguei lá, meu Deus. (CAJ 09)

(99) É, mas *eles* num me dão cobertô não, só me dá mais calça. Cobertô é difícil de *eles* me dá. (ITA 11)

A última variável selecionada pelo programa foi Localidade do falante:

Tabela 31- O pronome *eles* segundo a variável Localidade do falante

Localidade	Número de ocorrências/Total	Percentual	Peso relativo
Liberdade	147/509	28,9%	.407
Cajazeiras	141/495	23,1%	.514
Itapuã	204/627	29,2%	.564
Subúrbio	148/514	24%	.501
Total	640/2145	29,8%	--

Nível de significância: .017

Itapuã registra um claro indício de favorecimento de uso da estratégia *eles*, com peso relativo de .564, embora leve. Cajazeiras e Subúrbio se aproximam do ponto neutro, com .514 e .501, respectivamente, enquanto que Liberdade registra um desfavorecimento leve, com peso relativo de .407.

4.3.3 Você

De acordo com os resultados fornecidos pelo GOLDVARB X, as variáveis selecionadas como estatisticamente relevantes para o pronome *você* foram as que estão dispostas na tabela a seguir, em ordem de relevância:

Tabela 32 - Variáveis selecionadas - *você*

1º Referência ao falante
2º Faixa etária
3º Nível de referencialidade do agente
4º Realização fonética do sujeito indeterminado
5º Estada fora da comunidade
6º Modo
7º Escolaridade
8º Localidade

Os resultados para a primeira variável selecionada pelo programa estão dispostos na tabela a seguir:

Tabela 33 - O pronome *você* segundo a variável Referência ao falante

Referência ao falante	Número de ocorrências/Total	Percentual	Peso relativo
Inclui o falante	371/1338	27,7%	.651
Não inclui o falante	11/807	1,4%	.262
Total	382/2145	17,8%	--

Nível de significância: .001

Quando o falante é incluído na referência do sujeito indeterminado, o peso relativo da estratégia *você* é de .651, demonstrando favorecimento moderado. Quando o falante não é incluído, o peso relativo é de .262, indicando desfavorecimento moderado. Isso mostra que o pronome *você* é preferencialmente utilizado quando o falante pretende se incluir no contexto ao qual ele se refere, o que deixa clara a mudança semântica pela qual este pronome está passando. Apesar de as gramáticas tradicionais ainda tratarem o *você* como pronome de tratamento, ele já está implementado na fala como pronome pessoal, deixando, inclusive, de ser exclusivo de segunda pessoa, já que ele é mais utilizado em contextos nos quais o falante se inclui.

A seguir, a estratégia *você* e variável faixa etária:

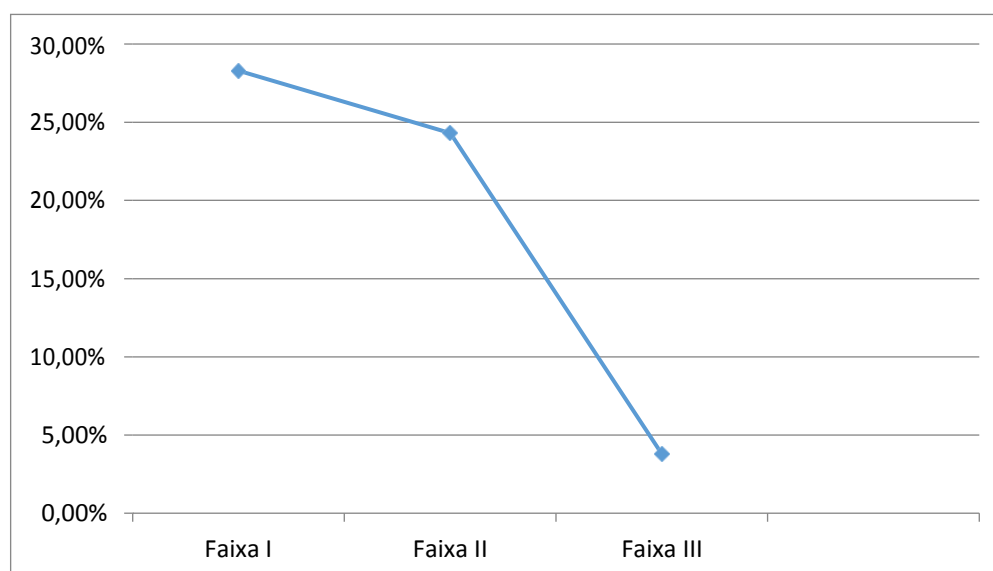
Tabela 34 - O pronome *você* segundo a variável Faixa etária

Faixa etária	Número de ocorrências/Total	Percentual	Peso relativo
Faixa 1	146/515	28,3%	.742
Faixa 2	206/849	24,3%	.629
Faixa 3	30/781	3,8%	.219
Total	382/2145	17,8%	--

Nível de significância: .001

A variável faixa etária apresentou resultados interessantes. De acordo com a tabela, podemos perceber que a faixa I, composta pelos informantes mais jovens, foi a que apresentou maior frequência de uso da estratégia *você*, com peso relativo de .742, bastante favorável. Em segundo, temos a faixa II, com peso relativo de .629, moderadamente favorável. A faixa etária que apresentou menor frequência foi a faixa III, composta pelos falantes mais velhos, com peso relativo de .219, desfavorável.

Esses resultados indicam que a estratégia indeterminadora *você* está entrando na comunidade através dos falantes mais jovens, que, em geral, são os que mais têm contato com o mundo externo à localidade e acabam por levar as inovações para o seio da comunidade de fala. Esses resultados ficam ainda mais claros no gráfico a seguir:

Gráfico 4 - O pronome *você* e a variável faixa etária

A próxima variável selecionada como numericamente relevante é o Nível de referencialidade do agente, e os resultados estão descritos a seguir:

Tabela 35 - O pronome *você* segundo a variável Nível de referencialidade do agente

Nível de referencialidade do Agente	Número de ocorrências/Total	Percentual	Peso relativo
Genérico	68/203	33,5%	.719
Parcial	306/1178	26%	.716
Indefinido	8/764	1%	.158
Total	382/2145	17,8%	--

Nível de significância: .001

Dentre os níveis de indeterminação analisados, podemos perceber que o que mais favorece o uso da estratégia *você* é o nível genérico, com peso relativo de .719. Em seguida, temos o nível de favorecimento intermediário ou parcial, com peso relativo de .716. Por último, temos o grau mais restrito, o chamado nível indefinido, com peso relativo de .158, bastante desfavorável.

A próxima variável selecionada é Realização fonética do sujeito indeterminado:

Tabela 36 - O pronome *você* segundo a variável Realização fonética do sujeito indeterminado

Realização fonética do sujeito indeterminado	Número de ocorrências/Total	Percentual	Peso relativo
Sujeito realizado	310/1119	27,7%	.624
Sujeito não realizado	72/1026	7%	.365
Total	382/2145	17,8%	--

Nível de significância: .001

Como apontam os resultados, a estratégia *você* é mais favorecida em contextos de realização fonética do pronome, com peso relativo de .624. Em contextos de sujeito não realizado, o peso relativo cai para .365, moderadamente desfavorável ao pronome em questão. Esse resultado confirma a hipótese de que há uma tendência de realização mais frequente dos pronomes indeterminadores, conforme já comentamos anteriormente.

A próxima variável selecionada para o pronome *você* foi Estada fora da comunidade:

Tabela 37- O pronome *você* segundo a variável Estada fora da comunidade

Estada fora da comunidade	Número de ocorrências/Total	Percentual	Peso relativo
Sim	111/479	23,2%	.645
Não	271/1666	16,3%	.457
Total	382/2145	17,8%	--

Nível de significância: .041

Os informantes que já moraram seis meses ou mais fora da comunidade são os que mais utilizam a estratégia *você*, com peso relativo de .645, moderadamente favorável. Para os informantes que nunca saíram da comunidade por mais de seis meses, o peso relativo é de .457, apenas levemente desfavorável à estratégia em questão. A seguir, os resultados para a variável Modo verbal:

Tabela 38 - O pronome *você* segundo a variável Modo

Modo	Número de ocorrências/Total	Percentual	Peso relativo
Realis	350/2077	16,9%	.490
Irrealis	32/68	47,1%	.778
Total	382/2145	17,8%	--

Nível de significância: .001

As frases em modo Irrealis, ou seja, aquelas em que o verbo aparece no modo subjuntivo, são as que mais favorecem o uso do pronome *você*, com peso relativo de .778. Em seguida, temos o modo Realis, equivalente ao modo indicativo, com peso relativo de .490, que se aproxima da neutralidade.

A penúltima variável selecionada como estatisticamente relevante para a estratégia *você* é Escolaridade.

Tabela 39 - O pronome *você* segundo a variável Escolaridade

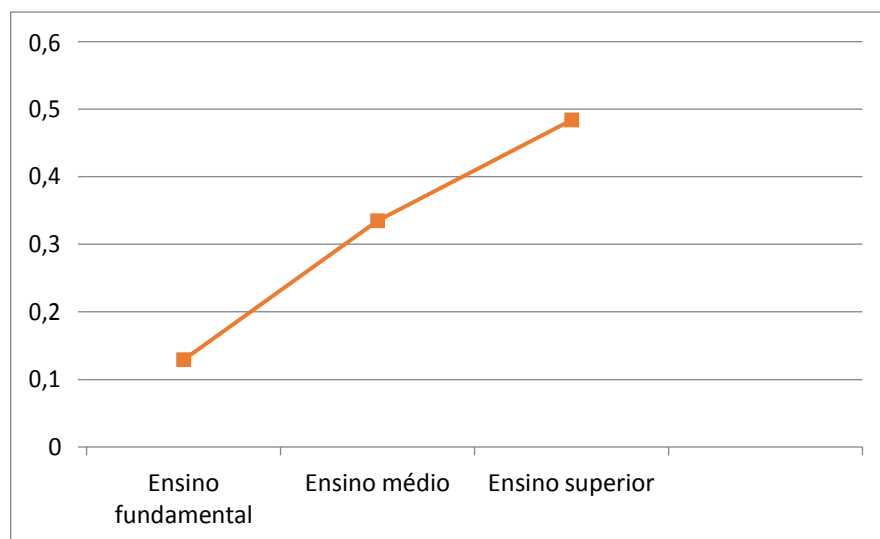
Escolaridade	Número de ocorrências/Total	Percentual	Peso relativo
Analfabeto	1/209	0,5%	.102
Semi-analfabeto	381/1936	19,7%	.558
Total	382/2145	17,8%	--

Nível de significância: .001

Conforme mostram os resultados, os informantes que mais utilizam a variante *você* são os semi-analfabetos, com .558 de peso relativo, indicando um favorecimento leve. A frequência para os informantes analfabetos é bem menor, com um peso relativo de desfavorecimento forte: .102. Esses resultados são bastante significativos se comparados aos encontrados por Carvalho (2010), que analisou o fenômeno e concluiu que quanto mais alto o nível de escolaridade do informante, maior a frequência de uso da estratégia *você*:

No caso da estratégia “você”, empregada como uma das marcas indeterminadoras do sujeito em Salvador, embora não seja tratada como uma das formas padrão abordadas pelas GT, pode ser considerada norma culta, uma vez que são os falantes com Ensino Superior completo que dominam o seu emprego. (p.140)

O gráfico a seguir mostra os resultados encontrados pelo autor:

Gráfico 5 - Uso da estratégia *você* em função da escolaridade (CARVALHO, 2008)

A variável Localidade foi a última a ser selecionada pelo programa:

Tabela 40 - O pronome *você* segundo a variável Localidade

Localidade	Número de ocorrências/Total	Percentual	Peso relativo
Liberdade	89/509	17,5%	.674
Cajazeiras	104/495	21%	.202
Itapuã	112/627	17,9%	.601
Subúrbio	77/514	15%	.527
Total	382/2145	17,8%	--

Nível de significância: .001

Dentre as localidades analisadas, a que apresentou maior frequência de uso do pronome *você* foi Liberdade, com peso relativo de .674. A segunda localidade com maior frequência foi Itapuã, com peso relativo de .601, seguida por Subúrbio e Cajazeiras, com respectivamente .527 e .202 de peso relativo.

4.3.4 Ø + V3PS

A quarta estratégia de maior ocorrência foi Ø + V3PS, que esteve presente em 316 ocorrências, o que corresponde a 14,7% do total. Para a rodada dessa variante, excluimos a variável Realização pronominal do sujeito, visto que essa variante apresenta lugar vazio para o sujeito pronominal, e também a variável desinência verbal, já que todas as ocorrências são de terceira pessoa do singular.

As variáveis selecionadas como estatisticamente relevantes foram as seguintes, em ordem de relevância:

Tabela 41 - Variáveis selecionadas - Ø + V3PS

- 1º Localidade
- 2º Faixa etária
- 3º Referência ao falante
- 4º Estada fora da comunidade

A seguir, os resultados encontrados para a variável Localidade:

Tabela 42 - A estratégia $\emptyset + V3PS$ segundo a variável Localidade

Localidade	Número de ocorrências/Total	Percentual	Peso relativo
Liberdade	138/509	27,1%	.741
Cajazeiras	77/495	15,6%	.404
Itapuã	52/627	8,3%	.421
Subúrbio	49/514	9,5%	.430
Total	316/2145	14,7%	--

Nível de significância: .001

Dentre as quatro localidades analisadas, a única que mais favorece o uso da estratégia $\emptyset + V3PS$ é Liberdade, com peso relativo de .741. As outras três localidades apresentaram resultados próximos, desfavorecendo a estratégia em questão: Subúrbio aparece em segundo lugar, com peso relativo de .430, seguida por Itapuã e Cajazeiras, com .421 e .404, respectivamente.

A próxima variável selecionada é Faixa etária, e os resultados apresentados pelo programa estão na tabela a seguir:

Tabela 43 - A estratégia $\emptyset + V3PS$ segundo a variável Faixa etária

Faixa etária	Número de ocorrências/Total	Percentual	Peso relativo
Faixa 1	89/515	17,3%	.581
Faixa 2	75/849	8,8%	.399
Faixa 3	152/781	19,5%	.557
Total	316/2145	14,7%	--

Nível de significância: .001

Como podemos ver na tabela acima, as únicas faixas etárias que favorecem o uso da estratégia indeterminadora $\emptyset + V3PS$ são a faixa I e a faixa III, com pesos relativos de .581 e .557, respectivamente. A faixa etária intermediária desfavorece medianamente o uso dessa estratégia, com peso relativo de .399.

A terceira variável selecionada foi Referência ao falante:

Tabela 44 - A estratégia \emptyset + V3PS segundo a variável Referência ao falante

Referência ao falante	Número de ocorrências/Total	Percentual	Peso relativo
Inclui o falante	160/1338	12%	.457
Não inclui o falante	156/807	19,5%	.570
Total	316/2145	14,7%	--

Nível de significância: .001

De acordo com os dados da tabela acima, podemos afirmar que os contextos nos quais o falante se inclui desfavorecem o uso da estratégia \emptyset + V3PS, com peso relativo de .457. Já os contextos em que o falante não se inclui favorecem o uso da estratégia \emptyset + V3PS, com peso relativo de .570.

A última variável selecionada para a estratégia \emptyset + V3PS foi Estada fora da comunidade:

Tabela 45 - A estratégia \emptyset + V3PS segundo a variável Estada fora da comunidade

Estada fora da comunidade	Número de ocorrências/Total	Percentual	Peso relativo
Sim	93/479	19,4%	.668
Não	223/1666	13,4%	.450
Total	316/2145	14,7%	--

Nível de significância: .001

Os informantes que nunca saíram da comunidade desfavorecem o uso da variante em questão, o que é indicado pelo peso relativo de .450. Há um favorecimento do uso por aqueles que já saíram da comunidade por mais de seis meses, com peso relativo de .668.

4.3.5 Nós

A última estratégia a ser analisada é o pronome *nós*, que foi a estratégia de menor ocorrência no *corpus*, com 3,9% de frequência (73 ocorrências). As variáveis selecionadas como estatisticamente relevantes para esta estratégia foram as seguintes:

Tabela 46 - Variáveis selecionadas – *nós*

- 1° Sexo
- 2° Localidade
- 3° Faixa etária
- 4° Realização fonética do sujeito indeterminado
- 5° Nível de referencialidade do agente
- 6° Tipo de frase
- 7° Escolaridade

A primeira variável a ser analisada é o Sexo, com os resultados na tabela abaixo:

Tabela 47 - O pronome *nós* segundo a variável Sexo

Sexo	Número de ocorrências/Total	Percentual	Peso relativo
Feminino	8/1025	0,8%	.280
Masculino	65/1120	5,8%	.704
Total	73/2145	3,4%	--

Nível de significância: .007

Como podemos ver acima, o pronome *nós* é preferencialmente utilizado pelos homens, com peso relativo de .704. Entre as mulheres, a frequência de uso do *nós* cai consideravelmente, e a análise registra um peso relativo de .280, bastante desfavorável.

A próxima variável selecionada foi Localidade do falante:

Tabela 48 - O pronome *nós* segundo a variável Localidade

Localidade	Número de ocorrências/Total	Percentual	Peso relativo
Liberdade	7/509	1,4%	.567
Cajazeiras	2/495	0,4%	.039
Itapuã	48/627	7,7%	.871
Subúrbio	16/514	3,1%	.618
Total	73/2145	3,4%	--

Nível de significância: .007

A localidade que mais favorece o uso do pronome *nós* é Itapuã, com peso relativo de .871. Em seguida, também favorecem essa estratégia Subúrbio, com peso relativo de .618, e Liberdade, com peso relativo de .567, enquanto que Cajazeiras, a desfavorece fortemente, com um peso relativo de .039.

A terceira variável selecionada foi a Faixa etária:

Tabela 49 - O pronome *nós* segundo a variável Faixa etária

Faixa etária	Número de ocorrências/Total	Percentual	Peso relativo
Faixa 1	5/515	1%	.128
Faixa 2	36/849	4,2%	.573
Faixa 3	32/781	4,1%	.717
Total	73/2145	3,4%	--

Nível de significância: .007

Os resultados da variável Faixa etária mostraram dados interessantes acerca do *status* da estratégia *nós* na comunidade. A variante *nós* é preferencialmente utilizada pelos falantes mais velhos, com peso relativo de .717, e pela faixa etária intermediária, com peso relativo de .573. Por último, a faixa etária I, composta pelos informantes mais jovens, desfavorece bastante essa estratégia, registrando um peso relativo de .128. Esses resultados nos mostram que a variante *nós* é preferencialmente utilizada pelos mais velhos e pouco utilizada pelos mais jovens, o que indica que a estratégia é mais conservadora e pode estar em vias de desaparecimento na comunidade em questão. Os resultados são similares aos encontrados por Ponte (2006) na análise do fenômeno no interior da Bahia, apresentados na tabela a seguir:

Tabela 50 - A estratégia *nós* e a variável Faixa etária no interior da Bahia – extraído de Ponte (2006)

Faixa etária	Número de ocorrências/Total	Percentual	Peso relativo
Faixa 1	72/1256	6%	.37
Faixa 2	104/1252	8%	.53
Faixa 3	96/786	12%	.66
Total	272/3294	8%	--

Em relação à variável Realização fonética do sujeito indeterminado, os resultados obtidos são os seguintes:

Tabela 51- O pronome *nós* segundo a variável Realização fonética do sujeito indeterminado

Realização fonética do sujeito indeterminado	Número de ocorrências/Total	Percentual	Peso relativo
Sujeito realizado	59/1119	5,3%	.613
Sujeito não realizado	14/1026	1,4%	.377
Total	73/2145	3,4%	--

Nível de significância: .007

Conforme já mencionamos anteriormente, os resultados obtidos mostram que em contexto de pronome sujeito *nós*, o sujeito pronominal foneticamente realizado é preferido, com um peso relativo de .613, contra um desfavorecimento mediano de .377 do sujeito não realizado.

A próxima variável selecionada pelo programa foi o Nível de referencialidade do agente, e os resultados estão dispostos a seguir:

Tabela 52 - O pronome *nós* segundo a variável Nível de referencialidade do agente

Nível de referencialidade do agente	Número de ocorrências/Total	Percentual	Peso relativo
Genérico	7/203	3,4%	.645
Parcial	59/1178	5%	.612
Indefinido	7/764	0,9%	.296
Total	73/2145	3,4%	--

Nível de significância: .007

Apenas dois níveis de referencialidade favorecem o pronome *nós*, sendo que o nível mais favorável, de favorecimento mediano, é o genérico, com peso relativo de .645 (100) e o nível intermediário, ou parcial, com peso relativo de .612 (101). Em contrapartida, o nível indefinido é desfavorável à estratégia em questão, com um peso relativo de .296 (102). Conforme já esperávamos, os resultados obtidos para essa variável são bastante semelhantes aos obtidos para a estratégia *a gente*, o que pode ser explicado pela proximidade semântica das duas estratégias, já que ambas são de primeira pessoa do plural.

(100) É esse o mundo que *nós* tamos vivendo! (LIB 01)

(101) *Nós* tomava banho de rio todo dia. (SUB 12)

(102) *Fomo* lá no delegado, o delegado não resolveu nada. (SUB 11)

A penúltima variável selecionada foi Tipo de frase:

Tabela 53 - O pronome *nós* segundo a variável Tipo de frase

Tipo de frase	Número de ocorrências/Total	Percentual	Peso relativo
Declarativa	71/1902	3,7%	.545
Negativa	2/239	0,8%	.189
Interrogativa	--	--	--
Total	73/2145	3,4%	--

Nível de significância: .007

As únicas frases que favorecem o uso do pronome *nós* são as declarativas, com peso relativo de .545 (103). As frases negativas desfavorecem bastante o uso do *nós*, com peso relativo de .189 (104).

(103) *Nós* aqui votamo nele e até hoje nada. (ITA 06)

(104) *Nós* num usa moto na pescaria. (ITA 02)

A última variável a ser analisada em relação ao pronome *nós* é Escolaridade:

Tabela 54 - O pronome *nós* segundo a variável Escolaridade

Escolaridade	Número de ocorrências/Total	Percentual	Peso relativo
Analfabeto	9/209	4,3%	.084
Semi-analfabeto	64/1936	3,3%	.564
Total	73/2145	3,4%	--

Nível de significância: .007

Os informantes que tiveram algum acesso à escolarização favorecem o uso da estratégia *nós*, com peso relativo de .564. Esse resultado já era esperado, já que a o pronome *nós* está mais relacionado à escolarização, por ser considerado uma variante padrão. Os informantes analfabetos desfavorecem fortemente o uso do pronome *nós*, apresentando um peso relativo de .084.

5 CONSIDERAÇÕES FINAIS

No início deste trabalho, estabelecemos como uma das principais metas entender como as circunstâncias sócio-históricas de formação de uma língua podem influenciar nas características estruturais dessa língua, analisando, no caso específico do Brasil, a influência das línguas indígenas e africanas para a formação do português popular brasileiro. Consideramos, então, que tais condições fizeram com que se pudesse observar, hoje, dois grandes padrões linguísticos no português brasileiro: a norma popular, oriunda da fala das camadas sociais menos favorecidas, e a fala culta, que derivou da fala das camadas mais favorecidas socialmente, cujos padrões de fala sempre estiveram voltados para os padrões portugueses.

Depois de discutir tais parâmetros sócio-históricos do português brasileiro, descrevemos, com o intuito de demonstrar a perspectiva de indeterminação do sujeito com a qual trabalharíamos, alguns estudos variacionistas já realizados com o mesmo tema tanto no português brasileiro quanto em outras variedades de língua portuguesa formadas através de contato, como o português moçambicano e o angolano. Discutimos conceitos relevantes, como os níveis de indeterminação do sujeito, a impessoalização e a relação entre o parâmetro do sujeito nulo e a indeterminação. Comentamos, também, a forma como o fenômeno é tratado na perspectiva da gramática tradicional, mostrando o quanto essa perspectiva é insuficiente, por não abordar todas as possibilidades que envolvem o fenômeno.

Depois de descrever detalhadamente o aporte teórico e metodológico no qual o trabalho seria realizado, partimos para a análise dos dados obtidos. Analisamos ocorrências com cinco estratégias pronominais: *a gente*, *nós*, *eles*, *você* e a estratégia $\emptyset + V3PS$, que consiste em uma estratégia verbal, na qual coloca-se o verbo na terceira pessoa do singular, sem sujeito exposto. Pretendíamos, inicialmente, analisar também a estratégia com o pronome indeterminador *se*, mas não encontramos nenhuma ocorrência dessa estratégia no *corpus*.

Codificamos um total de 2.145 ocorrências de sujeito indeterminado, sendo a estratégia *a gente* a mais frequente, com 34,2% do total de ocorrências. Em seguida, tivemos as estratégias *eles*, *você*, $\emptyset + V3PS$ e *nós*, com 29,8%, 17,8%, 14,7% e 3,4%, respectivamente. A maioria das estratégias encontradas foram realizadas foneticamente, o que indica uma possível mudança em direção aos sujeitos preenchidos que pode estar sendo

influenciada pela mudança no parâmetro do sujeito nulo (cf. seção 3). A baixa frequência da estratégia \emptyset + V3PS também é um indício dessa provável mudança.

Os resultados para a variável Desinência verbal também se mostram bastante significativos. Pretendíamos, ao analisar essa variável, verificar se há, no português popular se Salvador, um movimento em direção à redução da morfologia flexional, que se explicitaria através da alta frequência de verbos na terceira pessoa do singular. Essa hipótese se confirmou, pois a frequência de verbos na terceira pessoa do singular foi de 76%, estando estes verbos presentes em todas as estratégias de indeterminação. Podemos afirmar, então, que esse resultado está relacionado à alta frequência de sujeitos foneticamente realizados.

Alguns contextos linguísticos foram considerados estatisticamente relevantes para a análise. A variável linguística que se tornou mais relevante foi a Realização fonética do sujeito indeterminado, tendo sido selecionada por quatro estratégias indeterminadoras: *a gente*, *eles*, *você* e *nós*. O fato de essa variável linguística ter sido selecionada por quase todas as variáveis confirma a nossa hipótese: a de que os sujeitos foneticamente realizados são mais frequentes que os não realizados, movimento possivelmente provocado pela mudança na realização dos sujeitos de referência definida.

O nível de referencialidade do agente foi uma variável linguística selecionada para três estratégias indeterminadoras: *a gente*, *você* e *nós*. Em relação à estratégia *a gente*, há um favorecimento do uso para os níveis universal e genérico e um desfavorecimento para o nível específico, sendo que esse padrão se repete para a estratégia *nós*. Esses dados mostram que os recursos indeterminadores de primeira pessoa apresentam comportamento semelhante, como já esperávamos, e que essas estratégias representam os níveis mais abrangentes da indeterminação, já que incluem em seu escopo todas as pessoas do discurso.

A variável linguística Modo verbal foi selecionada como estatisticamente relevante para três estratégias indeterminadoras: *a gente*, *você* e *eles*. As estratégias *a gente* e *eles* foram favorecidas pelo modo Realis, e o modo Irrealis favoreceu o uso da estratégia *você*. Com isso, concluímos que a estratégia *você* é mais frequente em contextos hipotéticos, como os do exemplo a seguir:

(105) Se você for lá e não for morador do bairro, você não é atendido. (SUB 11)

A variável Tipo de frase foi selecionada por duas estratégias de indeterminação: *eles* e *nós*. Em ambos os casos, o uso das estratégias é favorecido em contextos de orações declarativas, que, devido à forma como as entrevistas são realizadas, são maioria no *corpus*.

Outra variável linguística selecionada por duas estratégias indeterminadoras foi Referência ao falante, tendo sido selecionada pelas variantes *você* e \emptyset + V3PS. A estratégia *você* é favorecida pelos contextos em que o falante é incluído na referência do sujeito indeterminado, o que mostra que essa é uma estratégia significativamente abrangente. O recurso indeterminador \emptyset + V3PS foi mais favorecido por contextos que não incluem o falante. O tipo de verbo e o paralelismo discursivo foram variáveis linguísticas que não demonstraram resultados numericamente relevantes.

Dentre as variáveis sociais analisadas, a que se mostrou mais relevante foi a Localidade do falante, tendo sido selecionada por todas as estratégias de indeterminação. Em relação aos perfis socioculturais dos habitantes desses bairros, observemos a tabela elaborada por Lima (2016, p.198–199):

Tabela 55 - Bairros de Salvador e características socioculturais

Bairro	Características socioculturais
Cajazeiras	Bairro novo, com a maioria dos habitantes nascidos no interior e com rede de relações tanto local quanto dispersa.
Itapuã	Bairro novo, com maioria dos habitantes nascidos no interior e com rede de relações dispersa.
Subúrbio	Bairro antigo, com maioria dos informantes nascidos no local e com rede de relações também local.
Liberdade	Bairro antigo, com maioria dos informantes nascidos no local e com rede de relações dispersa.

A distribuição das estratégias selecionadas por localidade ficou da seguinte forma: o pronome *a gente* foi favorecido pelo bairro Cajazeiras, o *eles*, pelo bairro Itapuã, o pronome *você* foi favorecido nos bairros Liberdade e Itapuã, a e estratégia $\emptyset + V3PS$ pelo bairro Liberdade, e o pronome *nós*, pelos bairros Itapuã e Subúrbio. Percebemos, por esses resultados, que mesmo os bairros com perfis socioculturais distintos selecionaram estratégias iguais, o que nos leva a concluir que essas diferenças não são significativas a ponto de traçar características sociolinguísticas diferentes entre os bairros.

A variável Sexo se mostrou relevante para três estratégias: *a gente*, *eles* e *nós*. Embora a tradição dos estudos sociolinguísticos afirme que as mulheres costumam apresentar comportamento mais conservador, não foi possível comprovar essa hipótese por estarmos tratando de um fenômeno que não sofre estigma. Sendo assim, os resultados para essa variável social não se mostraram muito significativos.

As variáveis sociais Faixa etária, Estada fora da comunidade e Escolaridade se mostraram relevantes para duas estratégias indeterminadoras. A Faixa etária se mostrou relevante para as estratégias *você* e *nós*, sendo as faixas mais jovens favoreceram o uso do *você*, e a faixa mais velha, favoreceu o uso do *nós*. Esses dados nos levam a inferir que a estratégia *você* é a mais inovadora na comunidade, já que está sendo introduzida pelas faixas etárias mais novas. Em relação ao *nós*, podemos afirmar que, conforme esperávamos, os resultados indicam que essa estratégia está caindo em desuso na comunidade, uma vez que é mais frequente na fala dos mais velhos e menos frequente entre os mais novos.

A variável Estada fora da comunidade se mostrou relevante para duas estratégias indeterminadoras: *você* e $\emptyset + V3PS$. O uso do pronome *você* foi favorecido por informantes que já moraram fora da comunidade por mais de seis meses, o que também confirma a hipótese de que essa é a estratégia inovadora, que está sendo introduzida na comunidade. A estratégia $\emptyset + V3PS$ também foi favorecida por informantes que já moraram fora da comunidade. Esse resultado vai de encontro ao que nós esperávamos, já que acreditávamos que essa variante seria resultado das não incorporações morfológicas características do processo de transmissão linguística irregular, hipótese que se confirmaria se essa variante fosse mais utilizada por informantes mais velhos e que nunca saíram da comunidade.

A última variável social a ser analisada é a Escolaridade. Essa variável também foi selecionada como relevante para duas estratégias de indeterminação: *você* e *nós*. Ambas as estratégias se mostraram favorecidas por informantes semi-analfabetos. Essa variável não se mostrou relevante para a estratégia $\emptyset + V3PS$, que, novamente, não nos permite relacionar essa estratégia ao o processo de Transmissão Linguística Irregular.

Ao fim dessa análise, pudemos perceber que as estratégias indeterminadoras se apresentam mais numerosas que as previstas nas gramáticas, o que nos leva a afirmar que fenômeno é amplo e passível de estudos mais detalhados, que relacionem a sua variabilidade às características socioculturais dos falantes. Percebemos, também, que há um distanciamento significativo entre os resultados obtidos na capital do estado e os obtidos nos estudos realizados no interior, o que nos leva a concluir que as grandes cidades têm sua própria dinâmica sociolinguística, que é marcada, sobretudo, pela aproximação entre os dialetos rurais e urbanos e pelo apagamento das marcas dialetais características do interior, processo que se materializa no conceito de dialeto *rurbano*.

Esperamos que esse estudo contribua para a compreensão da realidade sociolinguística do estado da Bahia e da cidade de Salvador, e que ele possa motivar outras análises sobre o tema em outros centros urbanos, abarcando, também, a dinâmica de funcionamento do fenômeno em outras normas linguísticas, tais como a norma culta.

REFERÊNCIAS

- ARAÚJO, Leonardo Eustáquio Siqueira. A variável faixa etária em estudos sociolinguísticos. *Estudos linguísticos*, v. 36, n. 2, p. 389-398, 2007.
- ARAÚJO, Silvana Silva de Farias. O uso variável da concordância verbal no português do Brasil (PB) e no português de Angola (PA): a história externa em foco. *13º COLÓQUIO DA LUSOFONIA E 5º ENCONTRO AÇORIANO*, Florianópolis. Atas. Florianópolis, 2010.
- ASSUNÇÃO, Janivam da Silva. *A indeterminação do sujeito na variedade linguística de Feira de Santana: um estudo variacionista*. Feira de Santana: UEFS (Dissertação de Mestrado), 2012.
- BAXTER, Alan Norman. Transmissão Geracional Irregular na História do Português Brasileiro - divergências nas vertentes afro-brasileiras. *Revista Internacional de Língua Portuguesa* v.14, p.72-90, 1995.
- BAXTER, Alan Norman; LUCCHESI, Dante. A transmissão linguística irregular. In. *O português afro-brasileiro*. Salvador: EDUFBA, p 101-124, 2009.
- BORTONI-RICARDO, Stella Maris. *Educação em língua materna: a sociolinguística na sala de aula*. São Paulo: Parábola, 2006.
- BORTONI-RICARDO, Stella Maris. *Manual de sociolinguística*. São Paulo: Contexto, 2014.
- BORTONI-RICARDO, Stella Maris. *Nós chegemu na escola, e agora?: sociolinguística & educação*. Parábola, 2005.
- BRITTO, Helena. Pronomes fracos nulos e lexicalizados: das línguas verdadeiramente pro-drop ao Português do Brasil (PB). *Cadernos de estudos linguísticos*, v. 34, 2011.
- CABREDO HOFHERR, P. *Arbitrary reading of third person pronominals*. In: Matthias Weisgerber (ed.). *Proceedings of the conference Sinn und Bedeutung 7 (sub 7) Arbeitspapiere des Fachreichs Sprachwissenschaften*, 2003.
- CALLOU, Dinah; BARBOSA, Afrânio; LOPES, Célia. O português do Brasil: polarização sociolinguística. In: CARDOSO, Suzana A. Marcelino; MOTA, Jacyra Andrade; MATTOS e SILVA, Rosa Virgínia (Orgs.). *Quinhentos anos de história linguística no Brasil*. Salvador: FUNCULTURAL, p. 259-291, 2006.
- CARDOSO, Suzana. *Geolinguística: tradição e modernidade*. São Paulo: Parábola, 2010.
- CASTRO, Yeda Pessoa de. A matriz africana no português do Brasil. In: CARDOSO, Suzana; MOTA, Jacyra; MATTOS e SILVA, Rosa Virgínia (Org.). *Quinhentos Anos de História Linguística do Brasil*. Salvador: Secretaria da Cultura e Turismo do Estado da Bahia, p. 81-116, 2006.

- CAVALCANTE, Silvia Regina de Oliveira. *O sujeito indeterminado na escrita padrão: a imprensa carioca nos séculos XIX e XX*. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, RJ, 1999.
- CHAMBERS, Jack; TRUDGILL, Peter. Estructura sociolingüística e inovación lingüística. In: *La dialectología*. Madrid: Visor Libros, 1994. p. 115-136.
- CHOMSKY, Noam. *Lectures on Government and Binding*. Dordrecht: Foris, 1981.
- CINTRA, Lindley; CUNHA, Celso. *Nova gramática do português contemporâneo*. 5 ed. Rio de Janeiro: Lexikon, 2008.
- COAN, Giovanna Ike. “Recebe (m)-se novidades”: construções com *se* em anúncios de revistas paulistanas. *Estudos Linguísticos*, v. 41, n. 2, p. 616-631, 2012.
- COSERIU, Eugenio. *Sincronia, diacronia e história: o problema da mudança lingüística*. Rio de Janeiro: Presença, 1979.
- CREISSELS, Denis. *Impersonal and related constructions: a typological approach*. 2008. Disponível em: <http://www.deniscreissels.fr/public/Creissels-impers.constr.pdf>. Acesso em 25 nov. 2016
- CYRANKA, Lúcia Furtado de Mendonça; NASCIMENTO, Livia Arcanjo do; RIBEIRO, Patricia Rafaela Otoni. A sociolingüística no ensino fundamental: resultados de uma pesquisa-ação. *Linhas Críticas*, v. 16, n. 31, p. 361-376, 2010.
- DIAS, Luiz Francisco; LADEIRA, Emiliana da Consolação. Domínio de referência na constituição do sujeito gramatical: o indefinido, o inespecífico e o indeterminado. *Revista (Con)textos Linguísticos*, v. 7, n. 8, p. 90-104, 2013.
- DIK, Simon. *The Theory of Functional Grammar. Part 2: Complex and Derived Constructions*. Berlin/New York: Mouton de Gruyter, 1997.
- DUARTE, Maria Eugênia Lamoglia. O sujeito de referência indeterminada em sentenças infinitivas. *Revista do GEL*, São José do Rio Preto, v. 5, n. 1, p. 9-30, 2008.
- DUARTE, Maria Eugênia Lamoglia. Sujeitos de referência definida e arbitrária: aspectos conservadores e inovadores na escrita padrão. *Revista Linguística*, Rio de Janeiro, v. 3, n. 1, p. 89-115, 2015.
- DUARTE, Maria Eugênia Lamoglia. *A perda do princípio “Evite Pronome” no PB*. Tese de Doutorado – UNICAMP, Campinas, 1995.
- DUARTE, Maria Eugenia Lamoglia; KATO, Mary; BARBOSA, Pilar. Sujeitos indeterminados em PE e PB. *CONGRESSO INTERNACIONAL DA ABRALIN*, p. 405- 409, 2001.
- FARACO, Carlos Alberto. *Norma culta brasileira: desatando alguns nós*. Parábola, 2008.
- FERREIRA, Mauro. *Aprender e praticar gramática*. São Paulo: FTD, 2003.

GAST, Volker. *Towards a typology of human impersonal pronouns*. Friedrich Schiller University Jena, 2015.

GODOY, Maria Alice Maschio. *A indeterminação do sujeito no interior paranaense: uma abordagem sociolinguística*. 128p. Dissertação (Mestrado em Letras e Linguística), Universidade Federal do Paraná, Paraná, 1999.

GUY, Gregory; ZILLES, Ana. *Sociolinguística quantitativa – Instrumental de análise*. São Paulo: Parábola, 2007.

KAISER, Georg. Sobre a (alegada) perda do sujeito nulo no português brasileiro. In: LOBO, Tânia et al. *Novos dados, novas análises: Tomo 1*. Salvador: EDUFBA, p. 11-42, 2006.

KITAGAWA, Chisato; LEHRER, Adrienne. Impersonal uses of personal pronouns. *Journal of pragmatics*, v. 14, n. 5, p. 739-759, 1990.

LEITE, Yonne; FRANCHETTO, Bruna. 500 anos de línguas indígenas no Brasil. In: *Quinhentos anos de história linguística do Brasil*. Salvador: Secretaria da Cultura e Turismo do Estado da Bahia, p. 16-61, 2006.

LIMA, Lanuza; ALMEIDA, Luanda. *Itapuã*. Disponível em: www.vertentes.ufba.br Acesso em: 15 de set. de 2016

LIMA, Rocha. *Gramática normativa da língua portuguesa*. Rio de Janeiro: José Olympio, 2013.

LOPES, Célia Regina dos Santos. Nós and a gente in standard spoken Brazilian Portuguese. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, v. 14, n. 2, p. 405-422, 1998.

LOPES, Célia Regina dos Santos. *A inserção de a gente no quadro pronominal do português*. Iberoamericana, 2003.

LUCCHESI, Dante. A deriva secular na formação do português brasileiro: uma visão crítica. In: LOBO, Tânia, CARNEIRO; Zenaide, SOLEDADÉ; Juliana, ALMEIDA, Ariadne; RIBEIRO, Silvana (orgs.). *Rosae: Linguística histórica, história das línguas e outras histórias*. Salvador: EDUFBA, 2012, p. 249-274

LUCCHESI, Dante. A realização do sujeito pronominal. In: *O português afro-brasileiro*. Salvador: EDUFBA, p. 167-184, 2009.

LUCCHESI, Dante. A simplificação morfológica na expressão do sujeito indeterminado no português afro-brasileiro. *Revista Linguística*, v. 10, n. 1, p. 277-298, 2015.

LUCCHESI, Dante. As duas grandes vertentes da história sociolinguística do Brasil. *D.E.L.T.A.*, São Paulo, v. 17, p. 97-130, 2001.

LUCCHESI, Dante. História do contato entre línguas no Brasil. **O português afrobrasileiro**. Salvador, EDUFBA, p. 41-73, 2009.

LUCCHESI, Dante. O conceito de transmissão linguística irregular e o processo de formação do português do Brasil. In: RONCARATI, Cláudia (org.). *Português brasileiro – contato linguístico, heterogeneidade e história*. Rio de Janeiro: Viveiros de Castro. p. 272-283, 2003.

LUCCHESI, Dante. *O conceito de transmissão linguística irregular*. Disponível em <http://www.vertentes.ufba.br/a-transmissao-linguistica-irregular>. Acesso em 27 de mai. de 2016

LUCCHESI, Dante. Parâmetros sociolinguísticos do português brasileiro. *Revista da ABRALIN*, v. 5, n. 1, p. 83-112, 2006.

LUCCHESI, Dante. Variação, mudança e norma: a questão brasileira. In: CARDOSO, Suzana. (org). *Diversidade linguística e ensino*. Salvador: EDUFBA p. 69-80, 1996.

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

LUCCHESI, Dante; BAXTER, Alan. Processos de crioulização na história sociolinguística do Brasil. In: *Quinhentos anos de história linguística do Brasil*. Salvador: Secretaria da Cultura e Turismo do Estado da Bahia, p. 163-218, 2006.

MACAMBYRA, Renata e BANDEIRA, Manuela. *Liberdade*. Disponível em: www.vertentes.ufba.br. Acesso em: 10 de jun. de 2010.

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

MATTHEIR, Klauss. *Sociolinguistics*. New York: Academic Press. p. 119-127, 1988.

MATTOS E SILVA, Rosa Virgínia. “*O português são dois...*”: *novas fronteiras, velhos problemas*. São Paulo: Parábola, 2004.

MATTOS E SILVA, Rosa Virgínia. Diversidade e unidade: a aventura linguística do português. *Revista Icalp*, v. 11, p. 60-72, 1988.

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

MENDES, Elisângela dos Passos. *Plataforma*. Disponível em: www.vertentes.ufba.br. Acesso em 10 de jun. de 2010.

MILANEZ, Wânia. *Recursos de indeterminação do sujeito*. Dissertação de mestrado. Faculdade de Letras, Unicamp, Campinas, 1982.

MINGAS, Amélia. *Interferência do kimbundu no português falado em Luanda*. Luanda: Caxinde, 2000.

MODESTO, Marcello. Sujeitos nulos em línguas de tópico proeminente. *Revista da ABRALIN*, v. 3, n. 1, p. 119-145, 2004.

MONTEIRO, José Lemos. *Para compreender Labov*. Petrópolis: Vozes, 2000.

NARO, Anthony Julius; SCHERRE, Maria Marta Pereira. *Origens do português brasileiro*. Parábola Books, 2007.

NEGRÃO, Esmeralda Vailat; VIOTTI, Evani. Estratégias de impessoalização no português brasileiro. In: FIORIN, José Luis; PETTER, Margarida. *África no Brasil. A formação da língua portuguesa*. São Paulo: Contexto, p.171-203, 2008.

NORTE, Gilberto Mariano; RIOS-NETO, Eduardo. *Línguas maternas e escolaridade em Moçambique*. Trabalho apresentado no XVI Encontro Nacional de Estudos Populacionais em Caxambú- MG – Brasil, 2008.

OLIVEIRA, Vítor Mateus Santos de. *A expressão do sujeito no português de Moçambique*. 109f. Dissertação de Mestrado – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2016.

PAIVA, Maria da Conceição de; DUARTE, Maria Eugenia Lamoglia (Orgs.). *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa, 2003.

PEREIRA, Deize Crespim. Uma análise funcionalista da indeterminação do sujeito no português popular falado em São Paulo. *Filologia e Linguística Portuguesa*. v. 15, n.2, p. 475-518, 2013.

PETTER, Margarida. Aspectos morfossintáticos comuns ao português angolano, brasileiro e moçambicano. *PAPIA-Revista Brasileira de Estudos Crioulos e Similares*, v. 19, n. 1, p. 201-220, 2009.

PONTE, Vanessa. *A indeterminação do sujeito no português popular do interior do estado da Bahia*. Salvador: UFBA. 127 p. Dissertação (Mestrado) - Programa de Pós-Graduação em Letras e Linguística, Universidade Federal da Bahia, 2008.

RODRIGUES, Ayrton. As línguas gerais sul americanas. *PAPIA-Revista Brasileira de Estudos Crioulos e Similares*, v. 4, n. 2, p. 6-18, 1996.

SANKOFF, David. Variable Rules. In: AMMON, Ulrich; DITTMAR, Norbert; SANKOFF, David; TAGLIAMONTE, Sali; SMITH, Eric. *Goldvarb X: a variable rule application for Macintosh and Windows*. 2005.

SANTOS, Jânio Ramos. *A cidade poli(multi)nucleada: A reestruturação do espaço urbano de Salvador*. Salvador: EDUFBA, 2013.

SCHERRE, Maria Marta Pereira. Paralelismo linguístico. *Revista de estudos da linguagem*, v. 7. n. 2, p. 29-59, 2012.

SETTI, Adriane Cristina Ribas. *A indeterminação do sujeito nas três capitais do Sul do Brasil*. 116 f. Dissertação (Mestrado)-Curso de Pós-graduação em Letras, Setor de Ciências Humanas, Letras e Artes, Universidade Federal do Paraná, Curitiba, 1997.

SIEWIERSKA, Anna. Ways of impersonalizing: pronominal vs. verbal strategies. In: MACKENZIE, Lachlan et al. *Language and cultures in contrast and comparison*. Amsterdam John Benjamins, 2007.

SILVA, Sylvio Bandeira de Mello. *Notas sobre o processo recente de urbanização/metropolização no estado da Bahia*. Salvador, 1, 1988. p. 31-52.

SILVEIRA, Vítor Cezário. *Estratégias de indeterminação em cartas do século XIX*. 2010. 98 f. Dissertação de Mestrado - Curso de Letras Vernáculas, UFRJ, Rio de Janeiro, 2010.

SOUZA, Elizete Maria de. *O uso do pronome 'eles' como recurso de indeterminação do sujeito*. Dissertação de mestrado, Belo Horizonte, 2007

SOUZA, Elizete Maria. *Sujeitos de referência arbitrária: uma classe homogênea?* Belo Horizonte: UFMG. 133p. Tese (Doutorado), 2013.

TEIXEIRA, Eliana Sandra Pitombo; ALMEIDA, Norma Lúcia Fernandes. A indeterminação do sujeito no português angolano: uma comparação com o português do Brasil. *PAPIA: Revista Brasileira de Estudos Crioulos e Similares*, v. 21, n. 1, p. 99-111, 2011.

TOURINHO, Cleber; ANTONINO, Vivian. *Cajazeiras*. Disponível em: www.vertentes.ufba.br. Acesso em 10 de jun. de 2010.

WEINREICH, Uriel; LABOV, William; HERZOG, Marvin. *Fundamentos empíricos para uma teoria da mudança linguística*. São Paulo: Parábola, 2006 [1975].

ANEXO A – CHAVE DE CODIFICAÇÃO

CHAVE DE CODIFICAÇÃO INDETERMINAÇÃO DO SUJEITO (somente ocorrências em tempo finito)

1. Variável dependente

(g) *a gente*/ Ø

(n) *nós* /-mos/ Ø

(v) *ocê*/ Ø

(e) *eles*/ Ø

(t) 3ª pessoa do singular/ Ø

(s) *se*

Variáveis explanatórias

2. Realização fonética do sujeito indeterminado

(+) Sujeito realizado

(-) Sujeito não realizado

3. Desinência Verbal

(3) 3ª Pessoa singular

(4) 1ª Pessoa plural

(6) 3ª Pessoa plural

4. Ocorrência anterior (por oração)

(P) Primeira menção

(g) *a gente* não-realizado

(G) *a gente* realizado

(n) *nós* não-realizado

(N) *nós* realizado

(v) *ocê* não-realizado

(V) *ocê* realizado

(L) *eles* realizado

- (m) *eles* não realizado
- (t) 3ª pessoa do singular
- (z) *se*

5. Referência ao falante

- (&) inclui o falante
- (%) não inclui o falante

6. Tipo de verbo

- (d) verbos transitivos *discendi* (dizer, falar, etc.)
- (c) verbos transitivos cognitivos (pensar, achar, etc.)
- (a) verbos transitivos de ação pontual (eu li um bom livro)
- (n) verbos transitivos de ação não pontual (eu leio muito)
- (e) verbos de ligação/estativos (*ser/estar* etc.)
- (i) verbos intransitivos inacusativos (nascer, adoecer, aparecer etc.)
- (t) verbos intransitivos (*correr, sorrir* etc.)
- (m) verbos de movimento (*ir, andar* etc.)
- (l) verbos locativos (*morar, viver*)
- (s) verbos aspectuais (*ele anda trabalhando muito*)
- (o) verbos modais (*deve.../tem que...*)
- (u) verbos de suporte/leve (*deu um grito*)
- (p) verbos de posse

7. Tipo de frase

- (I) interrogativa
- (D) declarativa
- (N) negativa

8. Nível de referencialidade do agente

- (g) genérico universal
- (p) grupo genérico
- (i) grupo específico cuja composição não é revelada ou é desconhecida

9. Modo

- (R) realis

(I) irrealis

Variáveis Sociais

11. Sexo

(F) feminino

(M) masculino

12. Faixa Etária

(1) faixa I

(2) faixa II

(3) faixa III

13. Estada fora da Comunidade

(*) sim

(#) não

14. Escolaridade

(@) Analfabeto

(\$) Semianalfabeto

15. Localidade

(C) Cajazeiras

(I) Itapuã

(L) Liberdade

(S) Subúrbio